NaturalProofs

Sean Welleck

Joint work with Gary Liu, Ronan Le Bras, Hanna Hajishirzi, Yejin Choi, Kyunghyun Cho

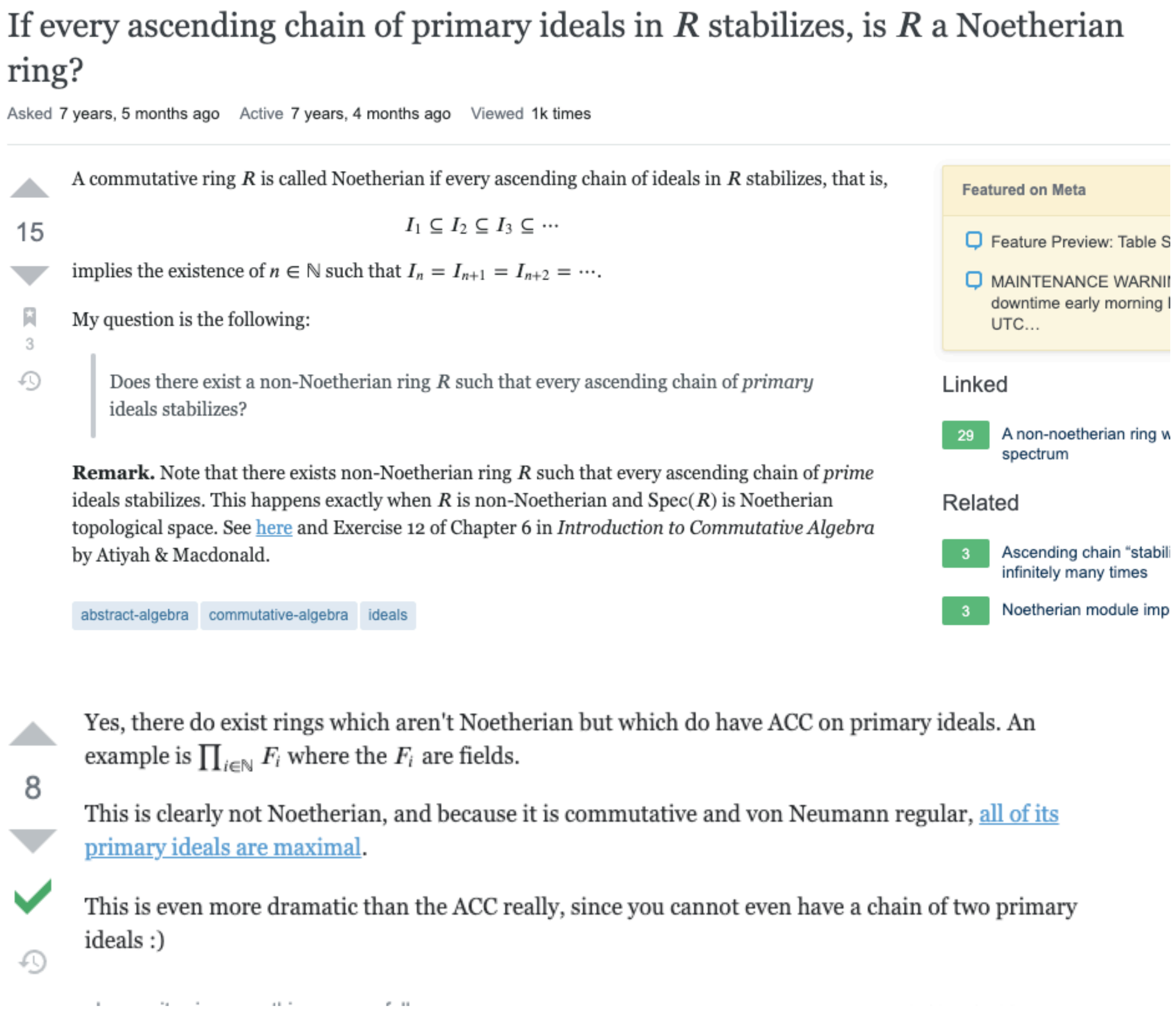
Overview

- **Motivation**: “Mathematical assistant”
- **Data**: Multi-domain NaturalProofs
- **Tasks**: Reference retrieval & generation
- **Future directions**

Motivation: guided learning

Mathematical assistant

- Proof-based mathematics is difficult to learn and self-study
  - Current approach requires human experts
- An interactive system capable of helping with arbitrary mathematics would require:
  - Mathematical reasoning ability
  - Natural language ability



If every ascending chain of primary ideals in R stabilizes, is R a Noetherian ring?

Asked 7 years, 5 months ago Active 7 years, 4 months ago Viewed 1k times

15

A commutative ring R is called Noetherian if every ascending chain of ideals in R stabilizes, that is,

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

implies the existence of $n \in \mathbb{N}$ such that $I_n = I_{n+1} = I_{n+2} = \cdots$.

My question is the following:

> Does there exist a non-Noetherian ring R such that every ascending chain of *primary* ideals stabilizes?

Remark. Note that there exists non-Noetherian ring R such that every ascending chain of *prime* ideals stabilizes. This happens exactly when R is non-Noetherian and $\mathrm{Spec}(R)$ is Noetherian topological space. See here and Exercise 12 of Chapter 6 in *Introduction to Commutative Algebra* by Atiyah & Macdonald.

8

Yes, there do exist rings which aren't Noetherian but which do have ACC on primary ideals. An example is $\prod_{i\in\mathbb{N}} F_i$ where the F_i are fields.

This is clearly not Noetherian, and because it is commutative and von Neumann regular, all of its primary ideals are maximal.

This is even more dramatic than the ACC really, since you cannot even have a chain of two primary ideals :)

Motivation: guided learning

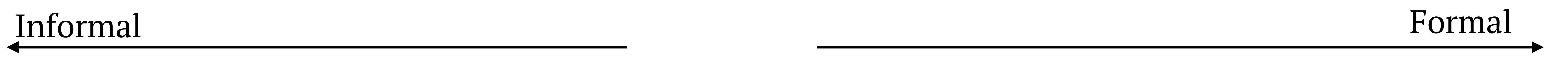
Informal — formal spectrum

- The mathematical assistant can be approached from a variety of angles

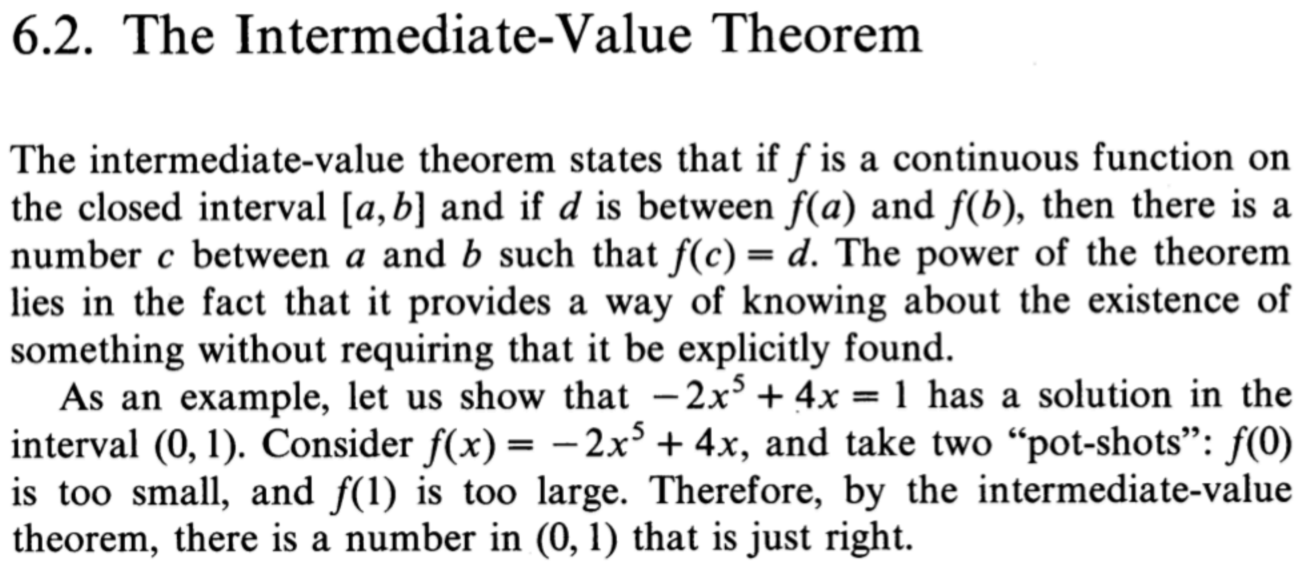
Motivation: guided learning

Informal — formal spectrum

- The mathematical assistant can be approached from a variety of angles



Textbook

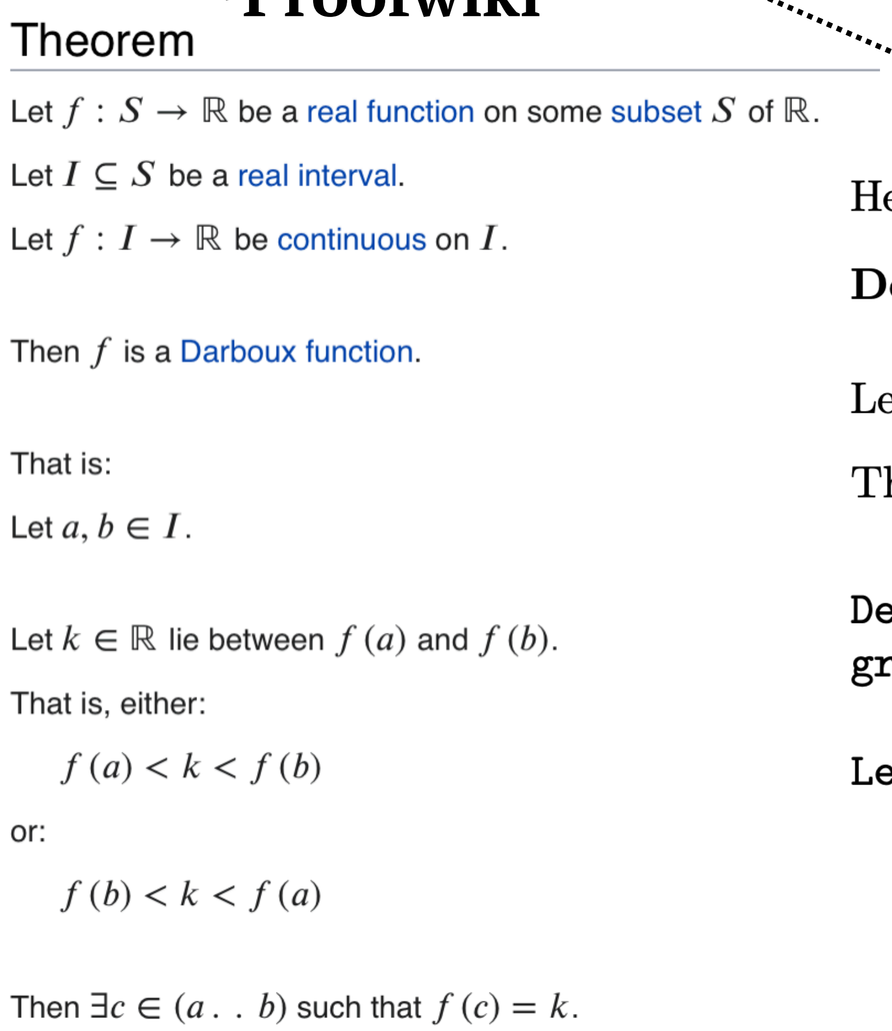


6.2. The Intermediate-Value Theorem

The intermediate-value theorem states that if f is a continuous function on the closed interval $[a, b]$ and if d is between $f(a)$ and $f(b)$, then there is a number c between a and b such that $f(c) = d$. The power of the theorem lies in the fact that it provides a way of knowing about the existence of something without requiring that it be explicitly found.

As an example, let us show that $-2x^5 + 4x = 1$ has a solution in the interval $(0, 1)$. Consider $f(x) = -2x^5 + 4x$, and take two "pot-shots": $f(0)$ is too small, and $f(1)$ is too large. Therefore, by the intermediate-value theorem, there is a number in $(0, 1)$ that is just right.

Proofwiki



Theorem

Let $f : S \to \mathbb{R}$ be a real function on some subset S of $\mathbb{R}$.

Let $I \subseteq S$ be a real interval.

Let $f : I \to \mathbb{R}$ be continuous on I.

Then f is a Darboux function.

That is:

Let $a, b \in I$.

Let $k \in \mathbb{R}$ lie between $f(a)$ and $f(b)$.

That is, either:

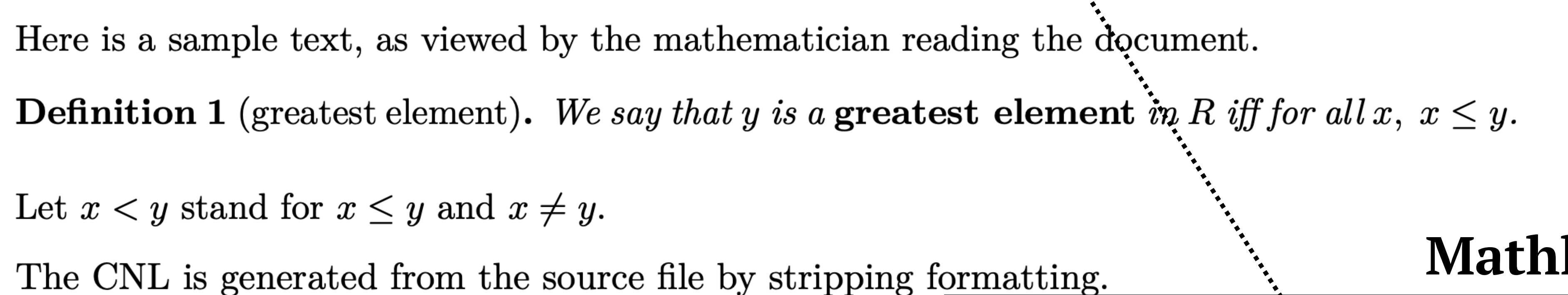
$$f(a) < k < f(b)$$

or:

$$f(b) < k < f(a)$$

Then $\exists c \in (a \,.\,.\, b)$ such that $f(c) = k$.

Colada
[Hales 2020]



Here is a sample text, as viewed by the mathematician reading the document.

Definition 1 (greatest element). *We say that y is a* **greatest element** *in R iff for all x, $x \leq y$.*

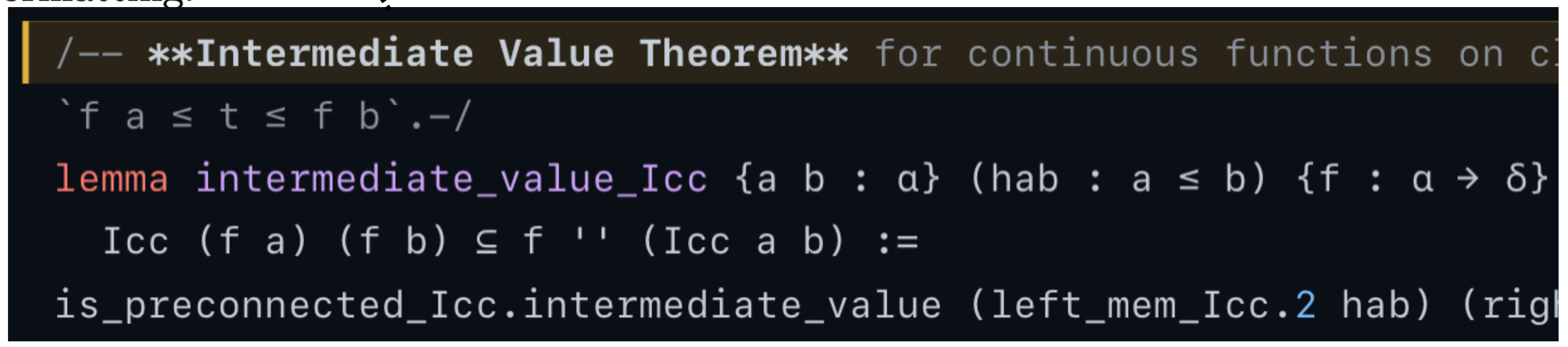
Let $x < y$ stand for $x \leq y$ and $x \neq y$.

The CNL is generated from the source file by stripping formatting.

```
Definition Label_greatest_element . We say that
greatestelement in R iff for all x , x \le y .

Let x < y stand for x \le y and x \ne y .
```

Mathlib/Lean

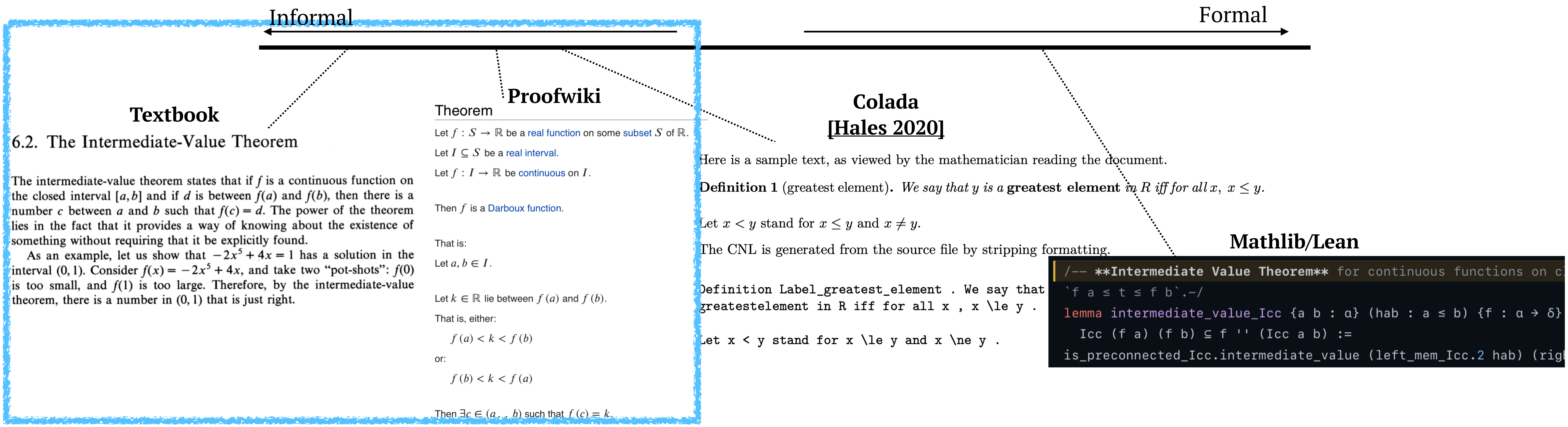


```
/-- **Intermediate Value Theorem** for continuous functions on c
`f a ≤ t ≤ f b`.-/
lemma intermediate_value_Icc {a b : α} (hab : a ≤ b) {f : α → δ}
  Icc (f a) (f b) ⊆ f '' (Icc a b) :=
is_preconnected_Icc.intermediate_value (left_mem_Icc.2 hab) (rig
```

Motivation: guided learning

Informal — formal spectrum

- The mathematical assistant can be approached from a variety of angles
- We consider the informal side here
    - Progress on/between any point of the spectrum is worthwhile



Textbook

6.2. The Intermediate-Value Theorem

The intermediate-value theorem states that if f is a continuous function on the closed interval $[a,b]$ and if d is between $f(a)$ and $f(b)$, then there is a number c between a and b such that $f(c) = d$. The power of the theorem lies in the fact that it provides a way of knowing about the existence of something without requiring that it be explicitly found.

As an example, let us show that $-2x^5 + 4x = 1$ has a solution in the interval $(0, 1)$. Consider $f(x) = -2x^5 + 4x$, and take two "pot-shots": $f(0)$ is too small, and $f(1)$ is too large. Therefore, by the intermediate-value theorem, there is a number in $(0, 1)$ that is just right.

Proofwiki

Theorem

Let $f : S \to \mathbb{R}$ be a real function on some subset S of $\mathbb{R}$.

Let $I \subseteq S$ be a real interval.

Let $f : I \to \mathbb{R}$ be continuous on I.

Then f is a Darboux function.

That is:

Let $a, b \in I$.

Let $k \in \mathbb{R}$ lie between $f(a)$ and $f(b)$.

That is, either:

$$f(a) < k < f(b)$$

or:

$$f(b) < k < f(a)$$

Then $\exists c \in (a \,.\,.\, b)$ such that $f(c) = k$.

Colada
[Hales 2020]

Here is a sample text, as viewed by the mathematician reading the document.

Definition 1 (greatest element)**.** *We say that y is a* **greatest element** *in R iff for all x, $x \le y$.*

Let $x < y$ stand for $x \le y$ and $x \ne y$.

The CNL is generated from the source file by stripping formatting.

```
Definition Label_greatest_element . We say that
greatestelement in R iff for all x , x \le y .

Let x < y stand for x \le y and x \ne y .
```

Mathlib/Lean

```
/-- **Intermediate Value Theorem** for continuous functions on c
`f a ≤ t ≤ f b`.-/
lemma intermediate_value_Icc {a b : α} (hab : a ≤ b) {f : α → δ}
  Icc (f a) (f b) ⊆ f '' (Icc a b) :=
is_preconnected_Icc.intermediate_value (left_mem_Icc.2 hab) (rig
```

Motivation: guided learning

Large-scale NLP methods for mathematics

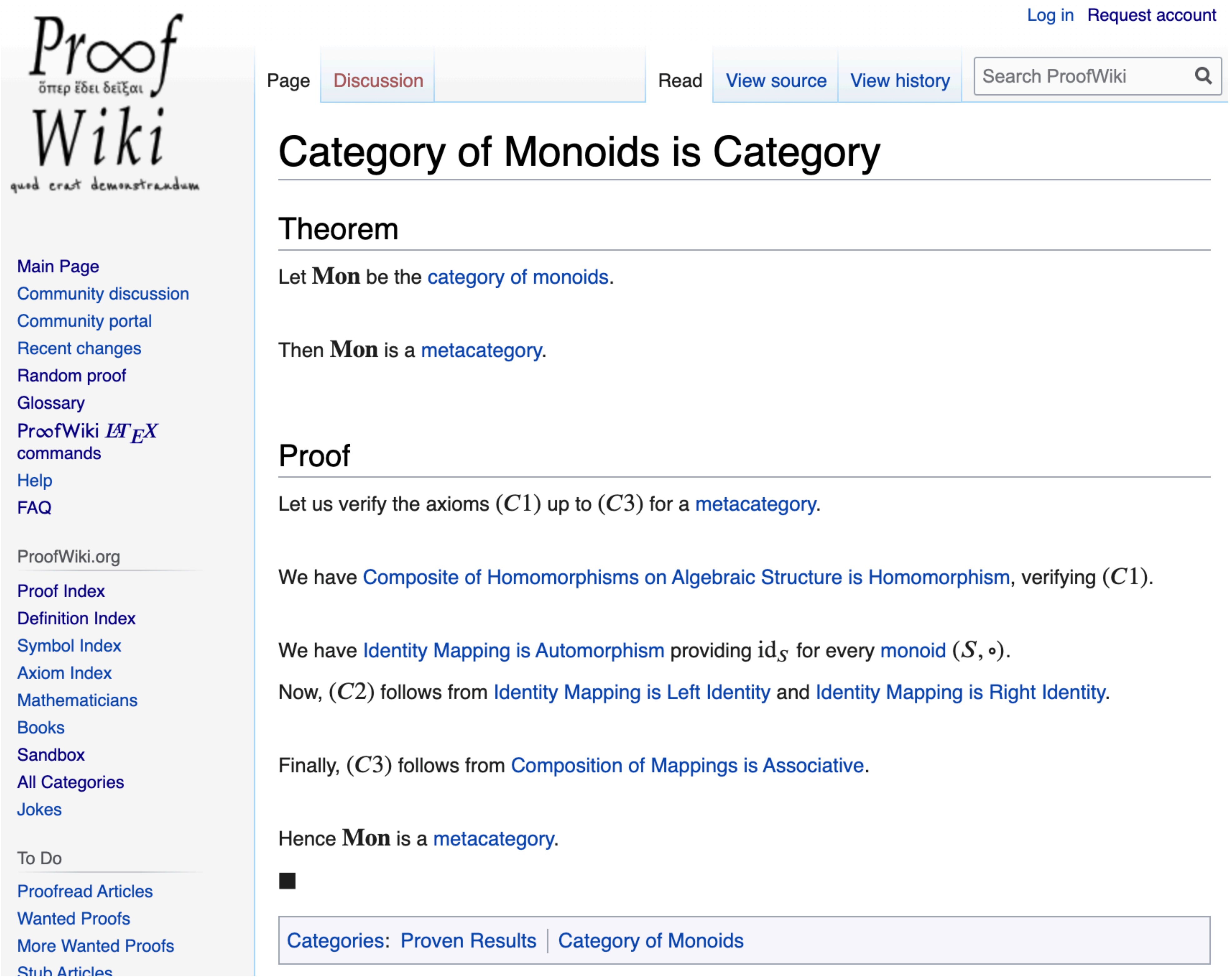
- Large-scale language models
(e.g. BERT [Devlin et al 2018], GPT [Radford et al 2019], T5 [Raffel et al 2019], BART [Lewis et al 2020])
    - Formal mathematics: e.g. GPT-f [Polu & Sutskever 2020], Skip-Tree [Rabe et al 2020], PACT [Han et al 2021]
    - Auto-formalization: e.g. [Kaliszyk et al 2014], [Wang et al 2020], [Szegedy 2020]
- Neural Retrieval
    - Dual encoder [Noguiera & Cho 2019], Dense Passage Retrieval [Karpukhin et al 2020]

- **Motivation**: “Mathematical assistant”
- **Data**: Multi-domain NaturalProofs
- **Tasks**: Reference retrieval & generation
- **Future directions**

Multi-domain NaturalProofs

Data sources

- **Broad-coverage mathematics**
  - **Proofwiki**: 20k theorems, 12.5k definitions



Category of Monoids is Category

Theorem

Let $\mathbf{Mon}$ be the category of monoids.

Then $\mathbf{Mon}$ is a metacategory.

Proof

Let us verify the axioms $(C1)$ up to $(C3)$ for a metacategory.

We have Composite of Homomorphisms on Algebraic Structure is Homomorphism, verifying $(C1)$.

We have Identity Mapping is Automorphism providing id_S for every monoid $(S, \circ)$.

Now, $(C2)$ follows from Identity Mapping is Left Identity and Identity Mapping is Right Identity.

Finally, $(C3)$ follows from Composition of Mappings is Associative.

Hence $\mathbf{Mon}$ is a metacategory.

$\blacksquare$

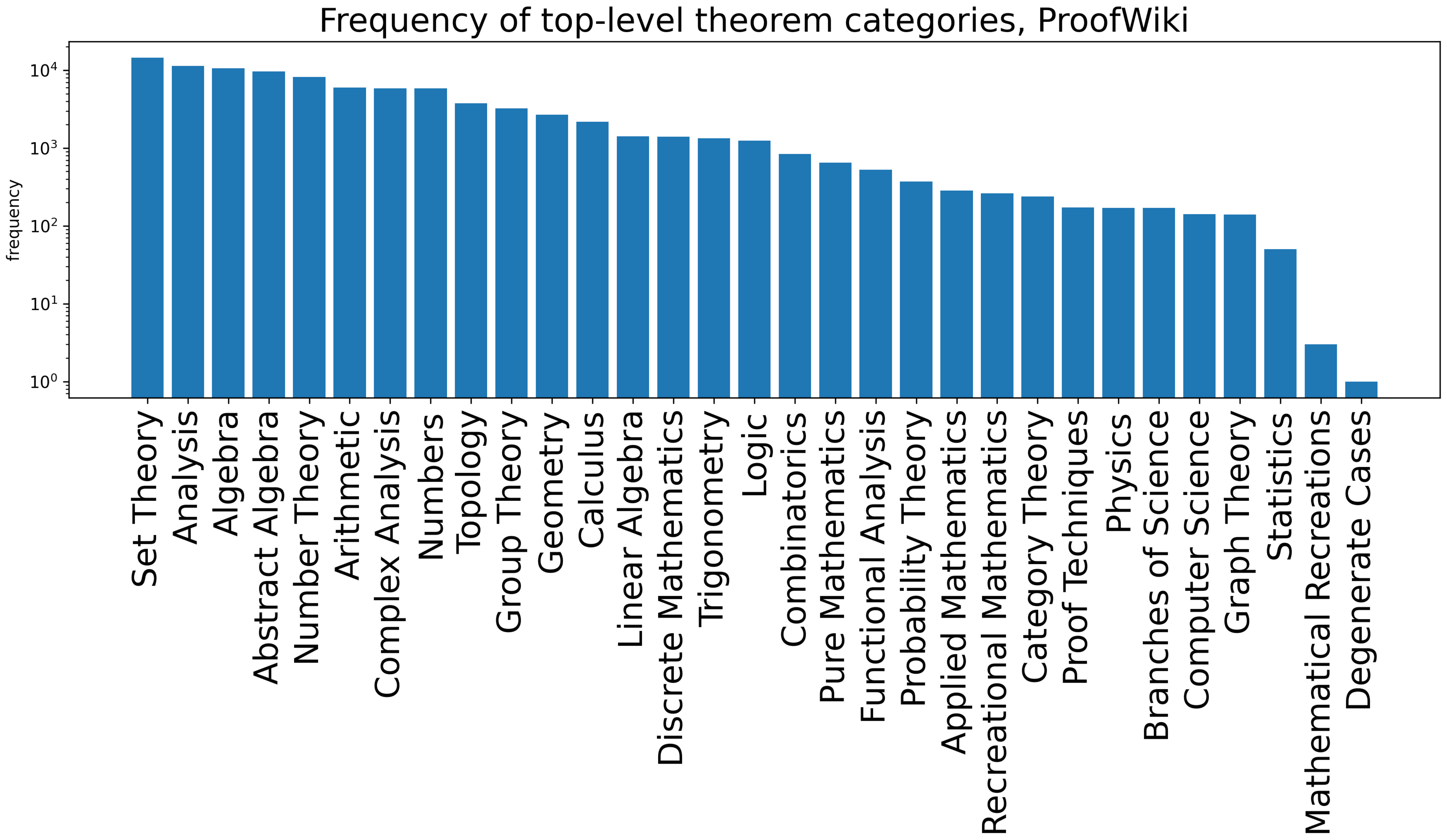
Categories: Proven Results | Category of Monoids

https://proofwiki.org/

Multi-domain NaturalProofs

Data sources

- **Broad-coverage mathematics**
  - **Proofwiki**: 20k theorems, 12.5k definitions
    - Large intersection with undergraduate curricula
    - Real-world users can benefit from tooling

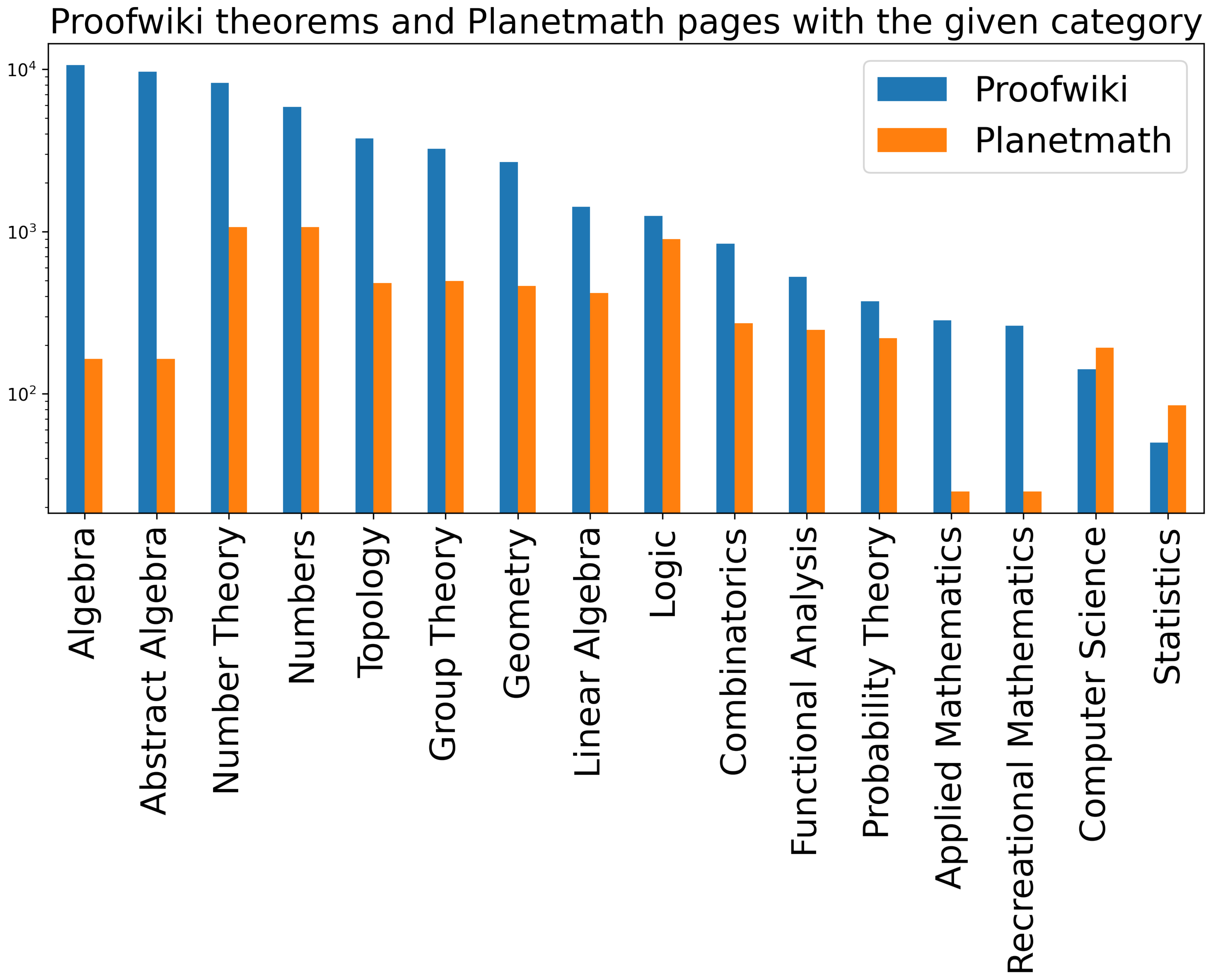


https://proofwiki.org/

Multi-domain NaturalProofs

Data sources

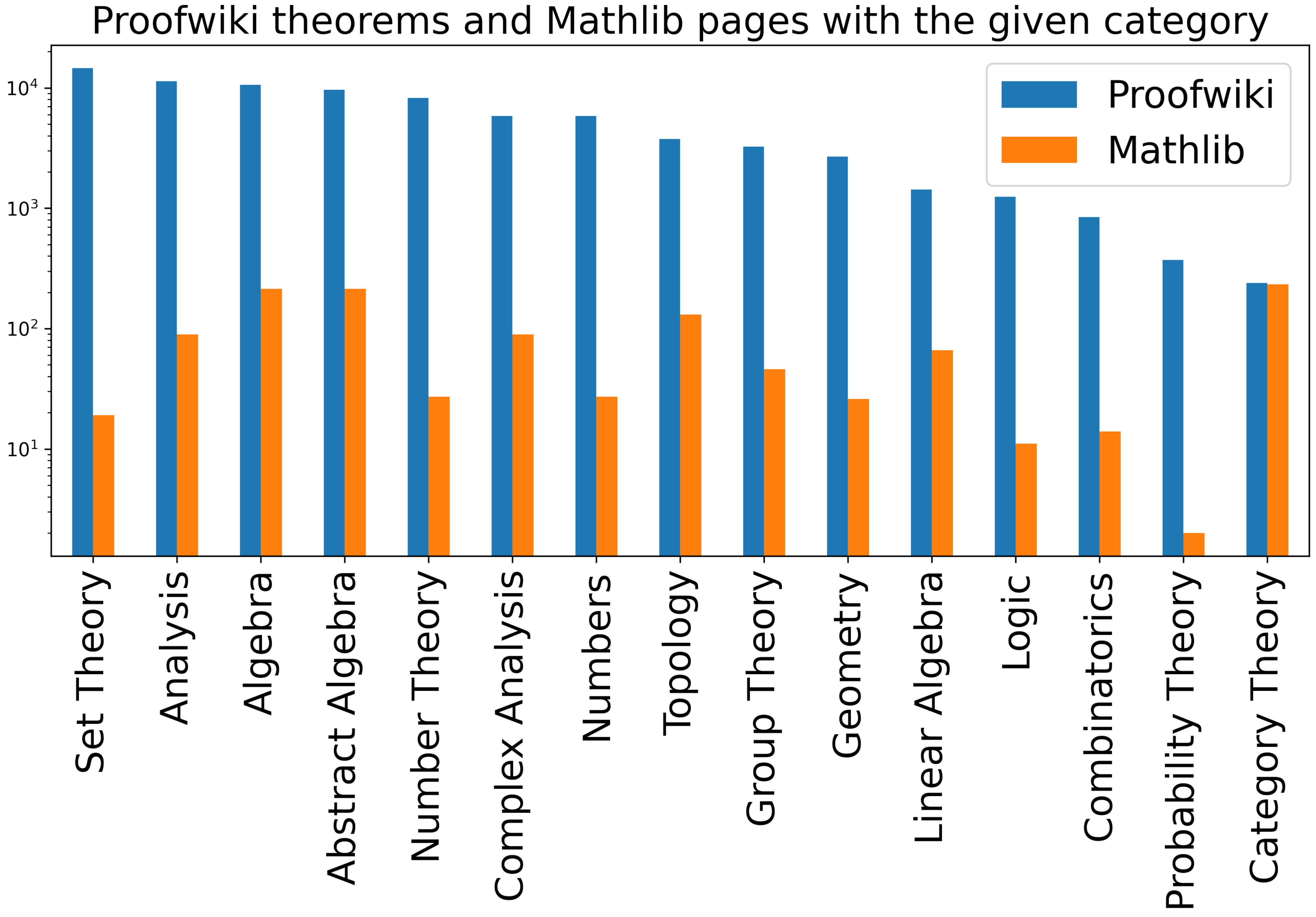
- **Broad-coverage mathematics**
  - **Proofwiki**: 20k theorems, 12.5k definitions
    - Large intersection with undergraduate curricula
    - Real-world users can benefit from tooling
    - Overlap: *PlanetMath*



Multi-domain NaturalProofs

Data sources

- **Broad-coverage mathematics**
  - **Proofwiki**: 20k theorems, 12.5k definitions
    - Large intersection with undergraduate curricula
    - Real-world users can benefit from tooling
    - Overlap: *PlanetMath*
    - Overlap: *Mathlib* (Lean)

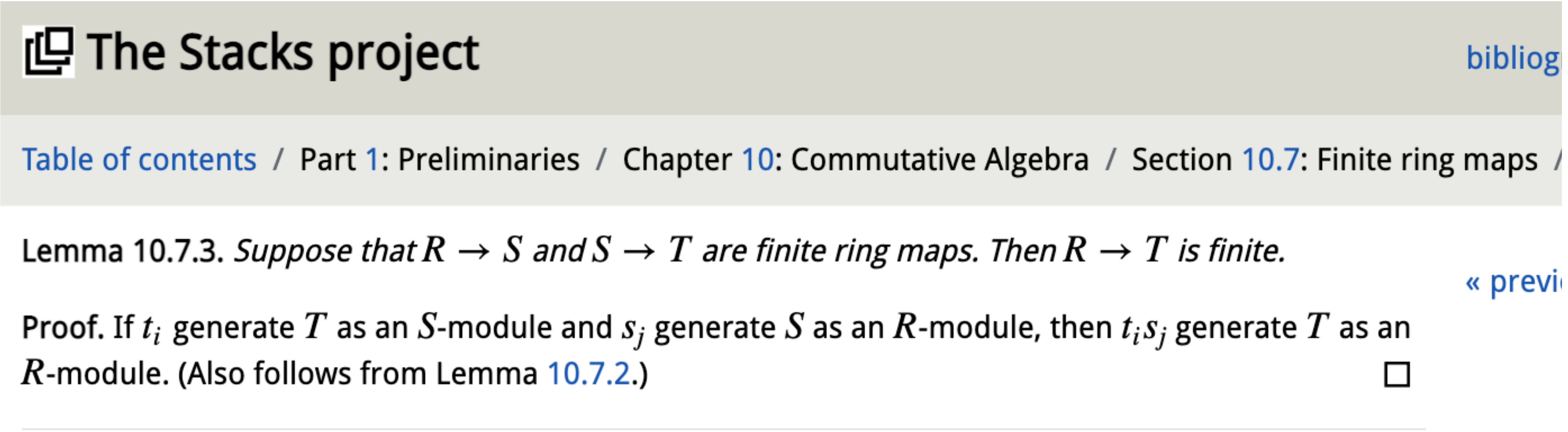


Multi-domain NaturalProofs

Data sources

- Broad-coverage

- **Deep-coverage**
  - **Stacks**: 12.5k theorems,
    1.7k definitions



The Stacks project

bibliog

Table of contents / Part 1: Preliminaries / Chapter 10: Commutative Algebra / Section 10.7: Finite ring maps /

« previ

Lemma 10.7.3. *Suppose that* $R \to S$ *and* $S \to T$ *are finite ring maps. Then* $R \to T$ *is finite.*

Proof. If t_i generate T as an S-module and s_j generate S as an R-module, then $t_i s_j$ generate T as an R-module. (Also follows from Lemma 10.7.2.) $\square$

https://stacks.math.columbia.edu/

Multi-domain NaturalProofs

Data sources

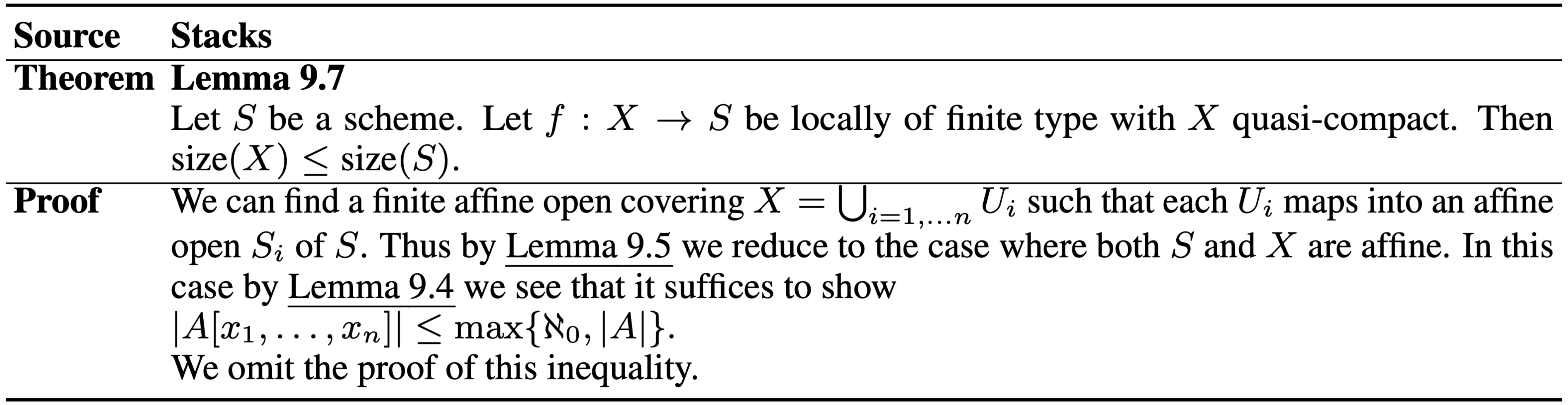
- Broad-coverage

- **Deep-coverage**

  - **Stacks**: 12.5k theorems,
    1.7k definitions

    - Research-level

    - Large-scale, density can benefit from search



Source	**Stacks**
Theorem	**Lemma 9.7** Let S be a scheme. Let $f : X \to S$ be locally of finite type with X quasi-compact. Then size$(X) \leq$ size(S).
Proof	We can find a finite affine open covering $X = \bigcup_{i=1,\ldots n} U_i$ such that each U_i maps into an affine open S_i of S. Thus by Lemma 9.5 we reduce to the case where both S and X are affine. In this case by Lemma 9.4 we see that it suffices to show $\lvert A[x_1, \ldots, x_n] \rvert \leq \max\{\aleph_0, \lvert A \rvert\}$. We omit the proof of this inequality.

Multi-domain NaturalProofs

Data sources

- Broad-coverage

- **Deep-coverage**

    - **Stacks**: 12.5k theorems,
      1.7k definitions

        - Research-level

        - Large-scale, density can benefit from search

        - Overlap: *Subset formalized in Lean*

        - Overlap: *Arxiv*

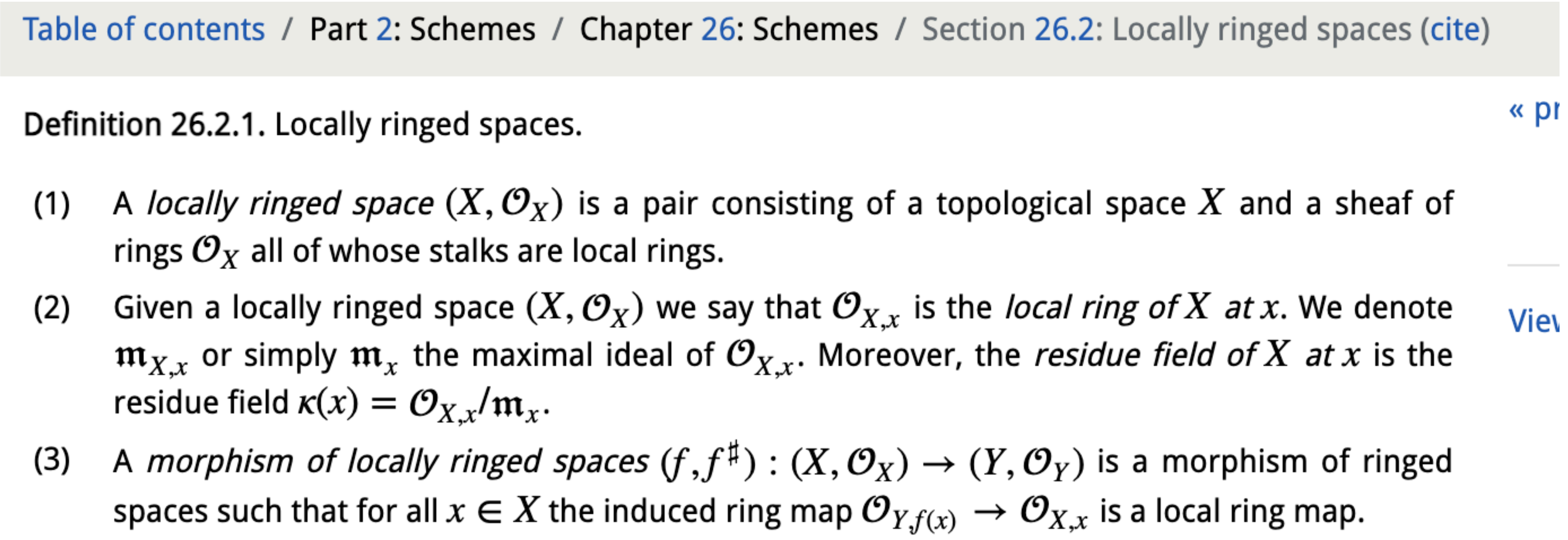


Table of contents / Part 2: Schemes / Chapter 26: Schemes / Section 26.2: Locally ringed spaces (cite)

« pr

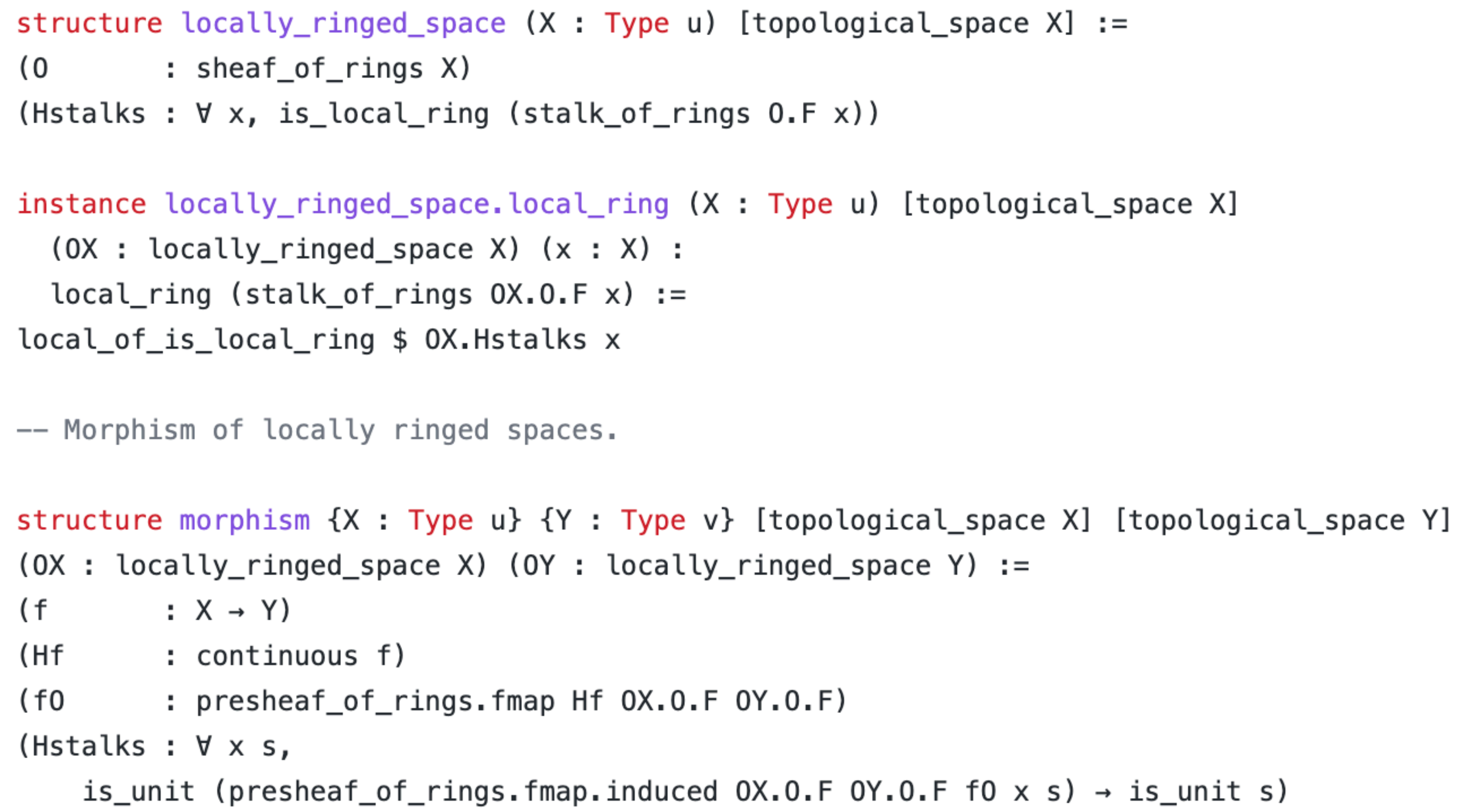
Vie

Definition 26.2.1. Locally ringed spaces.

(1) A *locally ringed space* $(X, \mathcal{O}_X)$ is a pair consisting of a topological space X and a sheaf of rings $\mathcal{O}_X$ all of whose stalks are local rings.

(2) Given a locally ringed space $(X, \mathcal{O}_X)$ we say that $\mathcal{O}_{X,x}$ is the *local ring of X at x*. We denote $\mathfrak{m}_{X,x}$ or simply $\mathfrak{m}_x$ the maximal ideal of $\mathcal{O}_{X,x}$. Moreover, the *residue field of X at x* is the residue field $\kappa(x) = \mathcal{O}_{X,x}/\mathfrak{m}_x$.

(3) A *morphism of locally ringed spaces* $(f, f^\sharp) : (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ is a morphism of ringed spaces such that for all $x \in X$ the induced ring map $\mathcal{O}_{Y,f(x)} \to \mathcal{O}_{X,x}$ is a local ring map.



```
structure locally_ringed_space (X : Type u) [topological_space X] :=
(O       : sheaf_of_rings X)
(Hstalks : ∀ x, is_local_ring (stalk_of_rings O.F x))

instance locally_ringed_space.local_ring (X : Type u) [topological_space X]
  (OX : locally_ringed_space X) (x : X) :
  local_ring (stalk_of_rings OX.O.F x) :=
local_of_is_local_ring $ OX.Hstalks x

-- Morphism of locally ringed spaces.

structure morphism {X : Type u} {Y : Type v} [topological_space X] [topological_space Y]
(OX : locally_ringed_space X) (OY : locally_ringed_space Y) :=
(f       : X → Y)
(Hf      : continuous f)
(fO      : presheaf_of_rings.fmap Hf OX.O.F OY.O.F)
(Hstalks : ∀ x s,
    is_unit (presheaf_of_rings.fmap.induced OX.O.F OY.O.F fO x s) → is_unit s)
```

Ramon Fernández Mir
https://github.com/ramonfmir/lean-scheme

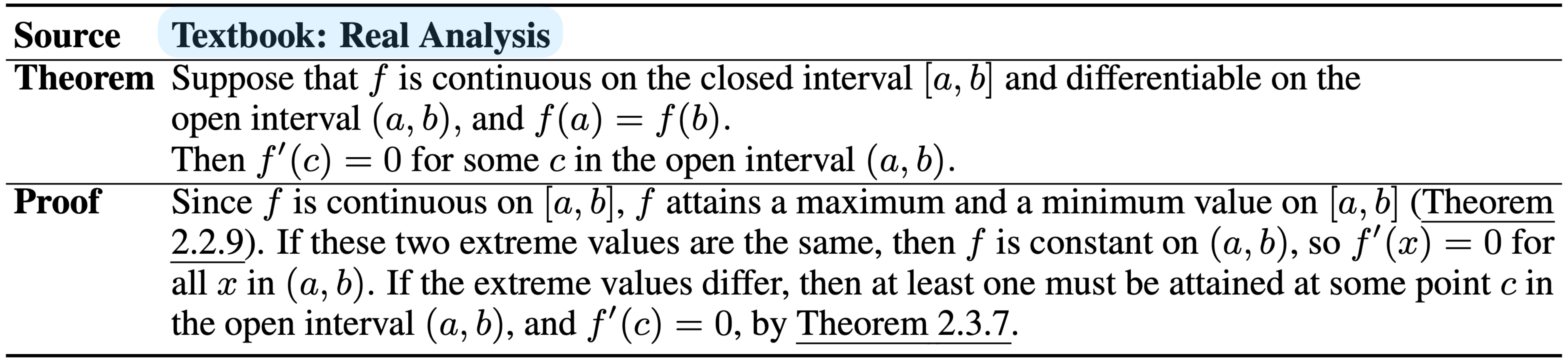
Multi-domain NaturalProofs

Data sources

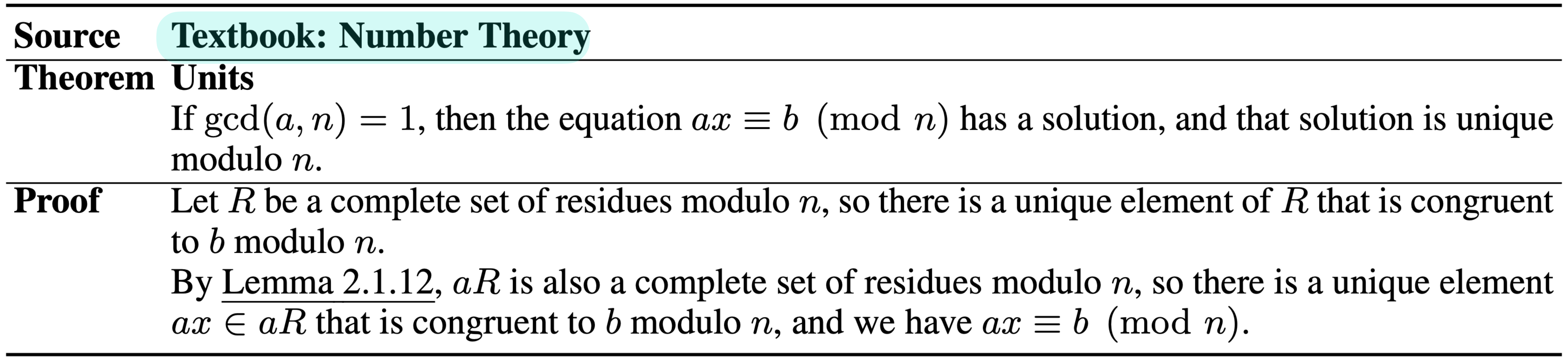
- Broad-coverage
- Deep-coverage
- **Low-resource**
  - **Textbooks**: Real Analysis, Number Theory
    - Education applications
    - ML challenge: OOD generalization

      298 / 69 theorems

      86 / 37 defintions



Source	**Textbook: Real Analysis**
Theorem	Suppose that f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), and $f(a) = f(b)$. Then $f'(c) = 0$ for some c in the open interval (a, b).
Proof	Since f is continuous on $[a, b]$, f attains a maximum and a minimum value on $[a, b]$ (Theorem 2.2.9). If these two extreme values are the same, then f is constant on (a, b), so $f'(x) = 0$ for all x in (a, b). If the extreme values differ, then at least one must be attained at some point c in the open interval (a, b), and $f'(c) = 0$, by Theorem 2.3.7.

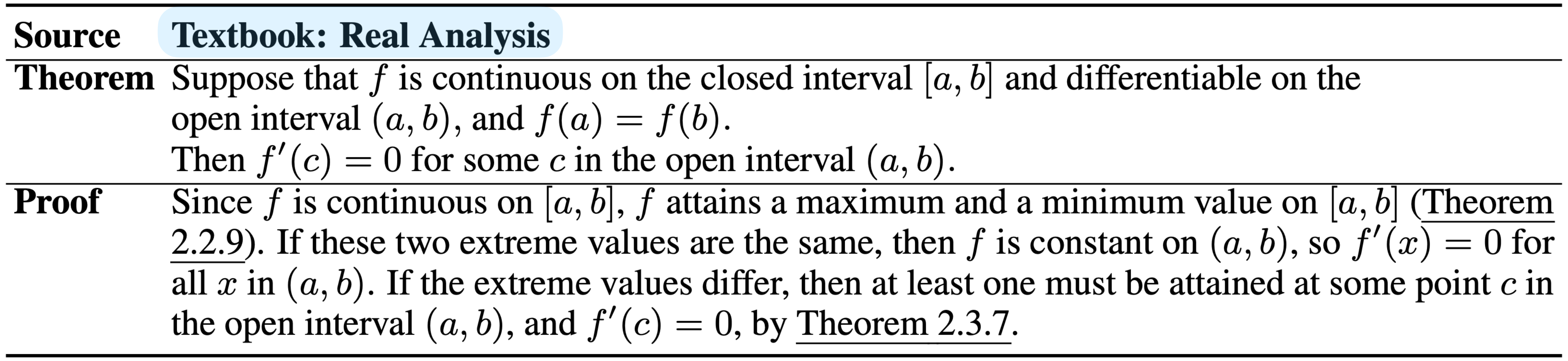


Source	**Textbook: Number Theory**
Theorem	**Units** If $\gcd(a, n) = 1$, then the equation $ax \equiv b \pmod{n}$ has a solution, and that solution is unique modulo n.
Proof	Let R be a complete set of residues modulo n, so there is a unique element of R that is congruent to b modulo n. By Lemma 2.1.12, aR is also a complete set of residues modulo n, so there is a unique element $ax \in aR$ that is congruent to b modulo n, and we have $ax \equiv b \pmod{n}$.

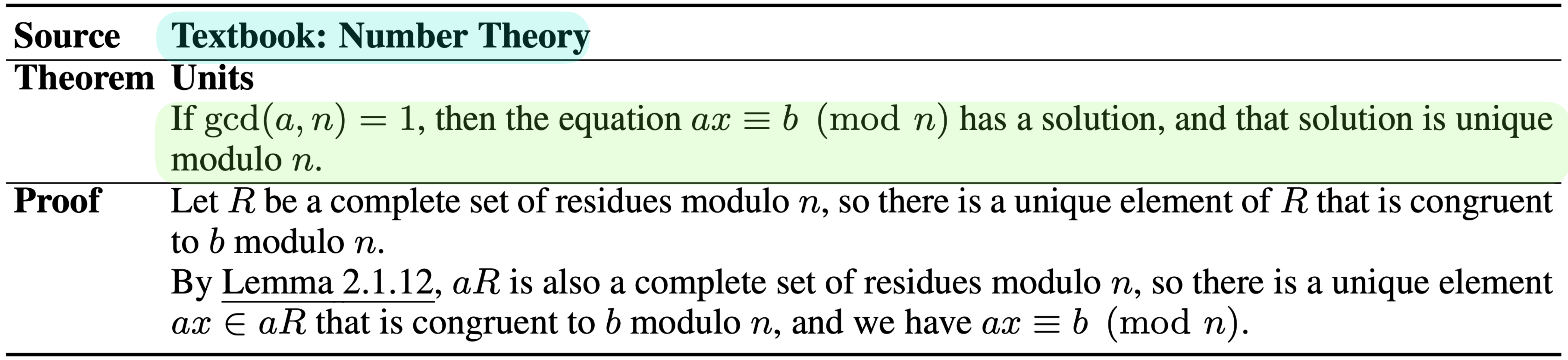
Multi-domain NaturalProofs

Data sources

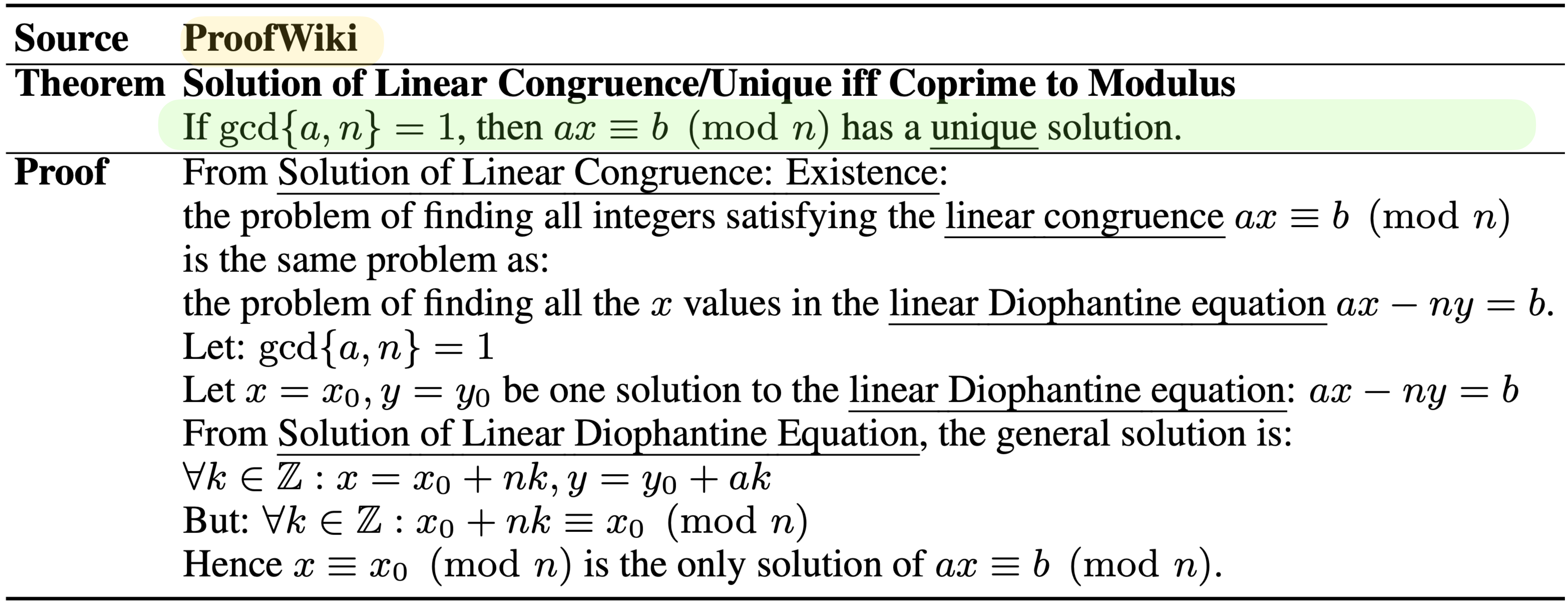
- Broad-coverage
- Deep-coverage
- **Low-resource**
  - **Textbooks**: Real Analysis, Number Theory
  - Education applications
  - ML challenge: OOD generalization
    298 / 69 theorems
    86 / 37 defintions
  - Overlap: **Proofwiki**



Source	**Textbook: Real Analysis**
Theorem	Suppose that f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), and $f(a) = f(b)$. Then $f'(c) = 0$ for some c in the open interval (a, b).
Proof	Since f is continuous on $[a, b]$, f attains a maximum and a minimum value on $[a, b]$ (Theorem 2.2.9). If these two extreme values are the same, then f is constant on (a, b), so $f'(x) = 0$ for all x in (a, b). If the extreme values differ, then at least one must be attained at some point c in the open interval (a, b), and $f'(c) = 0$, by Theorem 2.3.7.



Source	**Textbook: Number Theory**
Theorem	**Units** If $\gcd(a, n) = 1$, then the equation $ax \equiv b \pmod{n}$ has a solution, and that solution is unique modulo n.
Proof	Let R be a complete set of residues modulo n, so there is a unique element of R that is congruent to b modulo n. By Lemma 2.1.12, aR is also a complete set of residues modulo n, so there is a unique element $ax \in aR$ that is congruent to b modulo n, and we have $ax \equiv b \pmod{n}$.



Source	**ProofWiki**
Theorem	**Solution of Linear Congruence/Unique iff Coprime to Modulus** If $\gcd\{a, n\} = 1$, then $ax \equiv b \pmod{n}$ has a unique solution.
Proof	From Solution of Linear Congruence: Existence: the problem of finding all integers satisfying the linear congruence $ax \equiv b \pmod{n}$ is the same problem as: the problem of finding all the x values in the linear Diophantine equation $ax - ny = b$. Let: $\gcd\{a, n\} = 1$ Let $x = x_0, y = y_0$ be one solution to the linear Diophantine equation: $ax - ny = b$ From Solution of Linear Diophantine Equation, the general solution is: $\forall k \in \mathbb{Z} : x = x_0 + nk, y = y_0 + ak$ But: $\forall k \in \mathbb{Z} : x_0 + nk \equiv x_0 \pmod{n}$ Hence $x \equiv x_0 \pmod{n}$ is the only solution of $ax \equiv b \pmod{n}$.

Multi-domain NaturalProofs

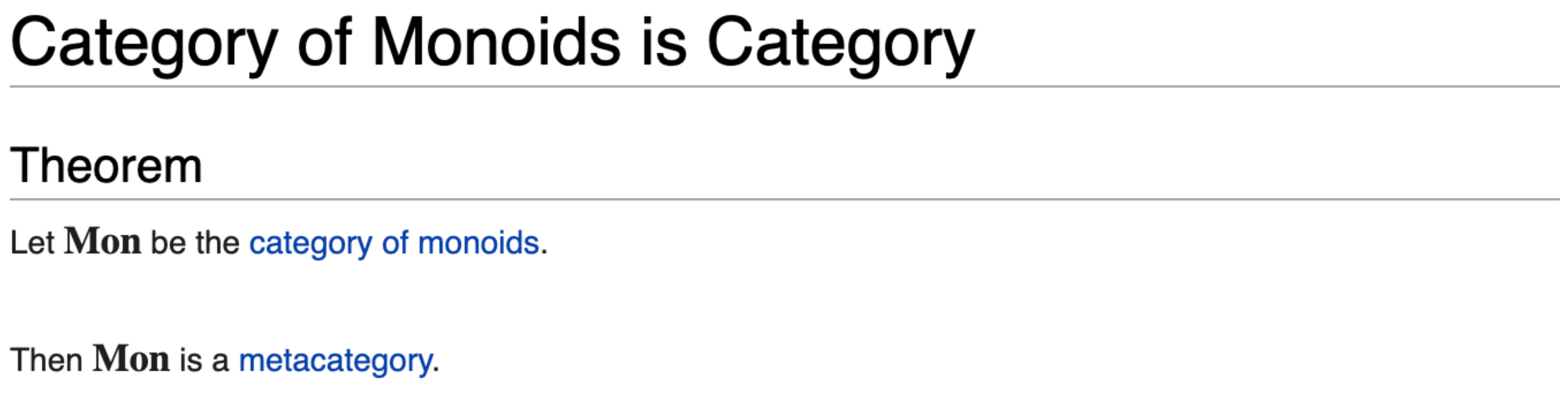
Schema

```
Dataset: {
  'dataset': {
    'theorems': [Statement],
    'definitions': [Statement],
    'others': [Statement],
Statement: {
  'id': int,
  'type': string,
  'label': string,
  'categories': [string],
  'toplevel_categories': [string], // ProofWiki only
  'recursive_categories': [string], // ProofWiki only
  'title': string,
  'contents': [string],
  'refs': [string],
  'ref_ids': [int],
  'proofs': [Proof], // for theorems only
}

Proof: {
  'contents': [string],
  'refs': [string],
  'ref_ids': [int],
}
```

Multi-domain NaturalProofs

Schema — Example



Category of Monoids is Category

Theorem

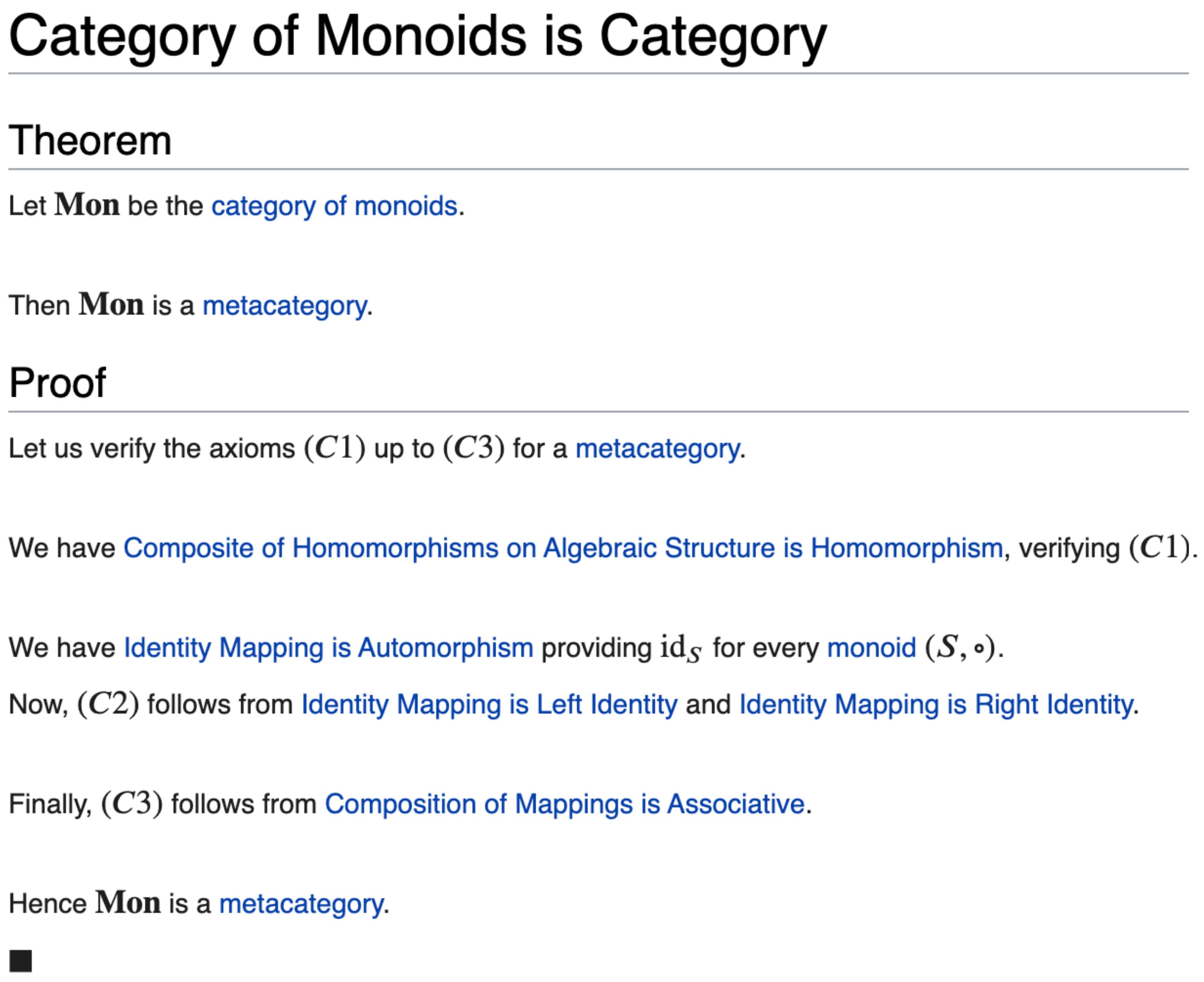
Let **Mon** be the category of monoids.

Then **Mon** is a metacategory.

```
{
  "id": 5480,
  "type": "theorem",
  "label": "Category of Monoids is Category",
  "title": "Category of Monoids is Category",
  "contents": [
    "Let $\\mathbf{Mon}$ be the [[Definition:Category of Monoids|category of monoids]].",
    "Then $\\mathbf{Mon}$ is a [[Definition:Metacategory|metacategory]]."
  ],
  "refs": [
    "Definition:Category of Monoids",
    "Definition:Metacategory"
  ],
  "ref_ids": [ 22919, 21454 ],
  "categories": [ "Category of Monoids" ],
  "toplevel_categories": [ "Algebra", "Set Theory", "Abstract Algebra", "Category Theory" ],
  "recursive_categories": [
    "Category Theory",
    "Algebra",
    "Abstract Algebra",
    "Category of Monoids",
    "Set Theory",
    "Examples of Categories"
  ],
```

Multi-domain NaturalProofs

Schema — Example



Category of Monoids is Category

Theorem

Let $\mathbf{Mon}$ be the category of monoids.

Then $\mathbf{Mon}$ is a metacategory.

Proof

Let us verify the axioms $(C1)$ up to $(C3)$ for a metacategory.

We have Composite of Homomorphisms on Algebraic Structure is Homomorphism, verifying $(C1)$.

We have Identity Mapping is Automorphism providing id_S for every monoid $(S, \circ)$.

Now, $(C2)$ follows from Identity Mapping is Left Identity and Identity Mapping is Right Identity.

Finally, $(C3)$ follows from Composition of Mappings is Associative.

Hence $\mathbf{Mon}$ is a metacategory.

■

```
"proofs": [
  {
    "contents": [
      "Let us verify the axioms $(C1)$ up to $(C3)$ for a [[Definition:Metacategory|metacategory]].",
      "We have [[Composite of Homomorphisms on Algebraic Structure is Homomorphism]], verifying $(C1)$.",
      "We have [[Identity Mapping is Automorphism]] providing $\\operatorname{id}_S$ for every
          [[Definition:Monoid|monoid]] $\\left({S, \\circ}\\right)$.",
      "Now, $(C2)$ follows from [[Identity Mapping is Left Identity]] and
          [[Identity Mapping is Right Identity]].",
      "Finally, $(C3)$ follows from [[Composition of Mappings is Associative]].",
      "Hence $\\mathbf{Mon}$ is a [[Definition:Metacategory|metacategory]].",
      "{{qed}}",

      "refs": [
        "Definition:Metacategory",
        "Composite of Homomorphisms is Homomorphism/Algebraic Structure",
        "Identity Mapping is Automorphism",
        "Definition:Monoid",
        "Identity Mapping is Left Identity",
        "Identity Mapping is Right Identity",
        "Composition of Mappings is Associative",
        "Definition:Metacategory"
      ],
      "ref_ids": [ 21454, 3852, 418, 19948, 217, 4387, 1494, 21454 ]
```

Multi-domain NaturalProofs

Summary

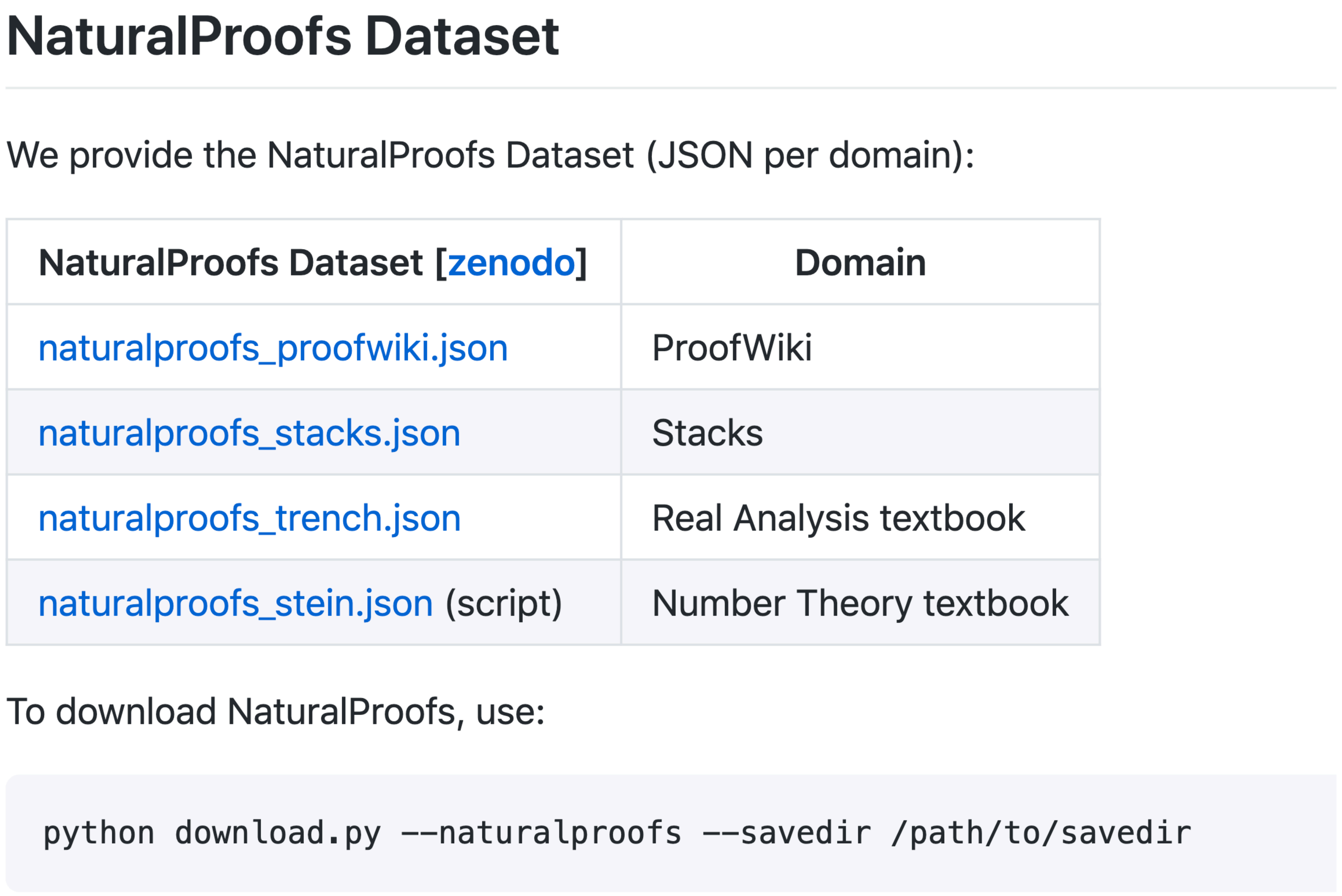
- **~30k theorems & proofs**
- **~14k definitions**
- **~2k other pages** (e.g. axioms)

	Source	All	PWiki	Stacks	RA	NT
Theorem	N	*32,579*	*19,734*	*12,479*	*298*	*68*
	Tokens	46.7	38.2	60.6	33.6	23.7
	Lines	5.9	3.6	9.7	8.4	4.5
	Refs	1.8	2.8	0.2	0.0	0.0
Proof	N	*32,012*	*19,234*	*12,479*	*235*	*64*
	Tokens	181.5	199.3	155.5	128.9	97.2
	Lines	24.9	25.8	23.4	36.1	16.1
	Refs	5.6	7.4	3.0	1.6	0.9
Definition	N	*14,230*	*12,420*	*1,687*	*86*	*37*
	Tokens	48.4	45.0	73.2	58.6	32.6
	Lines	5.0	4.2	10.7	13.3	5.1
	Refs	2.9	3.3	0.4	0.0	0.0
Other	N	*1,974*	*1,006*	*968*	–	–
	Tokens	212.1	286.1	135.2	–	–
	Lines	34.4	46.7	21.7	–	–
	Refs	5.7	9.2	2.0	–	–

Multi-domain NaturalProofs

Summary

- **~30k theorems & proofs**
- **~14k definitions**
- **~2k other pages** (e.g. axioms)



NaturalProofs Dataset

We provide the NaturalProofs Dataset (JSON per domain):

NaturalProofs Dataset [zenodo]	Domain
naturalproofs_proofwiki.json	ProofWiki
naturalproofs_stacks.json	Stacks
naturalproofs_trench.json	Real Analysis textbook
naturalproofs_stein.json (script)	Number Theory textbook

To download NaturalProofs, use:

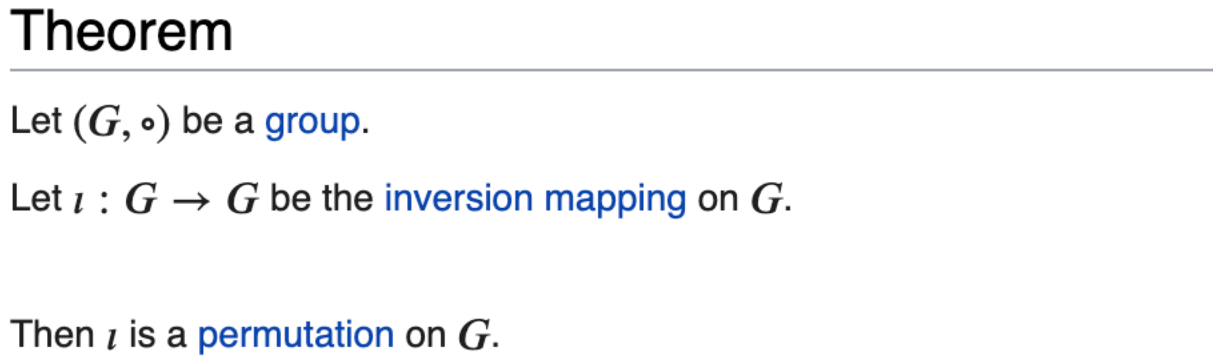
```
python download.py --naturalproofs --savedir /path/to/savedir
```

https://github.com/wellecks/naturalproofs

- **Motivation**: "Mathematical assistant"
- **Data**: Multi-domain NaturalProofs
- **Tasks**: Reference retrieval & generation
- **Future directions**

NaturalProofs tasks

Mathematical reasoning



Theorem

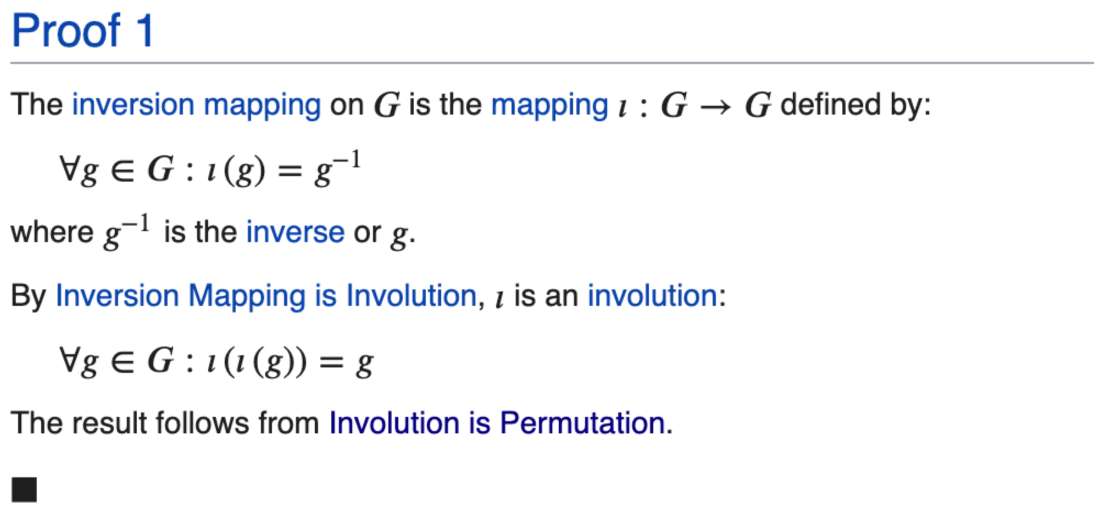
Let $(G, \circ)$ be a group.

Let $\iota : G \to G$ be the inversion mapping on G.

Then ι is a permutation on G.

Mathematical reasoner

David Hilbert



Proof 1

The inversion mapping on G is the mapping $\iota : G \to G$ defined by:

$$\forall g \in G : \iota(g) = g^{-1}$$

where g^{-1} is the inverse or g.

By Inversion Mapping is Involution, ι is an involution:

$$\forall g \in G : \iota(\iota(g)) = g$$

The result follows from Involution is Permutation.

■

NaturalProofs tasks

Mathematical reasoning

Theorem

Let $(G, \circ)$ be a group.

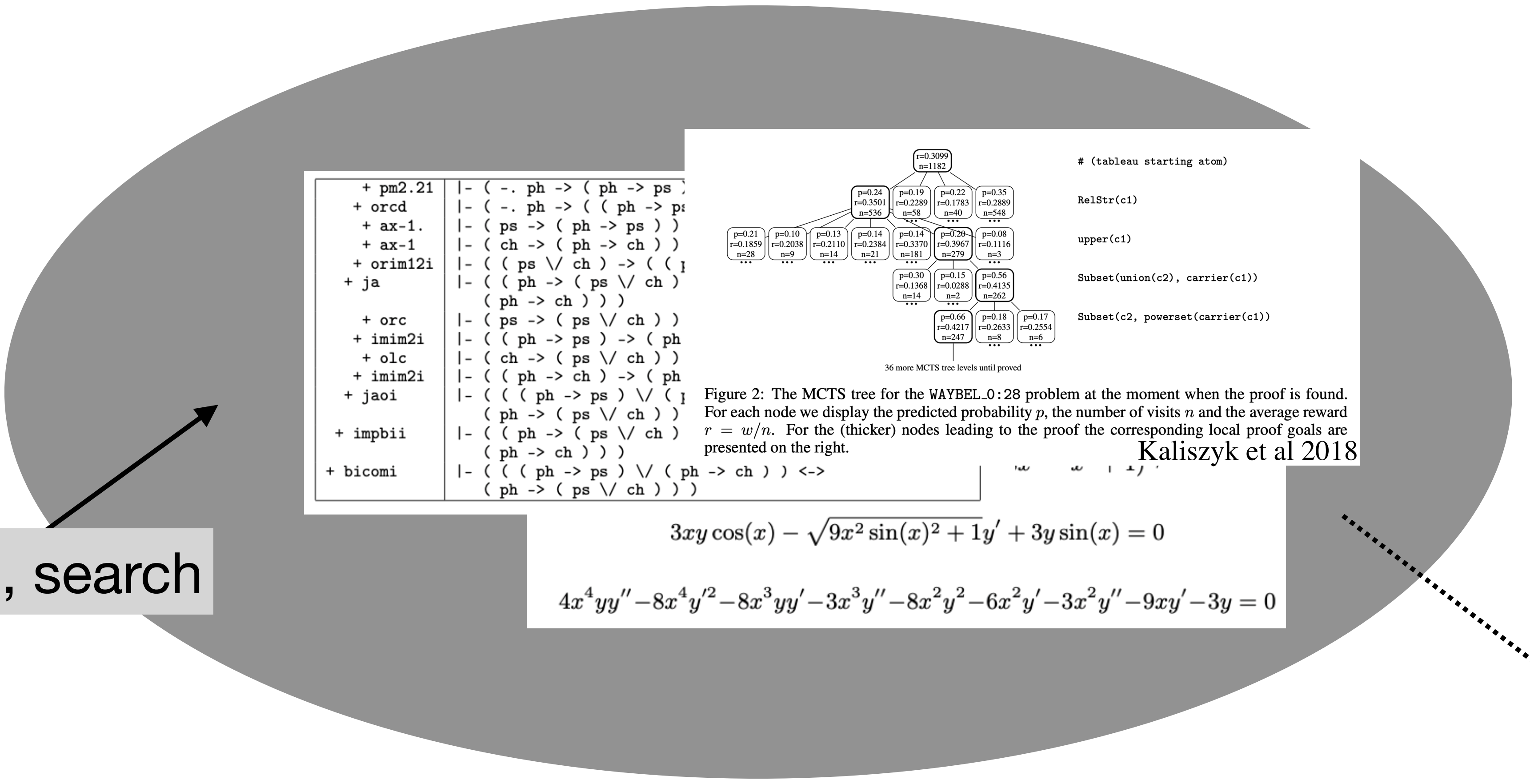
Let $\iota : G \to G$ be the inversion mapping on G.

Then ι is a permutation on G.

Mathematical reasoner

David Hilbert

Symbolic, search



```
  + pm2.21   |- ( -. ph -> ( ph -> ps
  + orcd     |- ( -. ph -> ( ( ph -> p
   + ax-1.   |- ( ps -> ( ph -> ps ) )
   + ax-1    |- ( ch -> ( ph -> ch ) )
  + orim12i  |- ( ( ps \/ ch ) -> ( (
 + ja        |- ( ( ph -> ( ps \/ ch )
                ( ph -> ch ) ) )
   + orc     |- ( ps -> ( ps \/ ch ) )
  + imim2i   |- ( ( ph -> ps ) -> ( ph
   + olc     |- ( ch -> ( ps \/ ch ) )
  + imim2i   |- ( ( ph -> ch ) -> ( ph
 + jaoi      |- ( ( ( ph -> ps ) \/ (
                ( ph -> ( ps \/ ch ) )
 + impbii    |- ( ( ph -> ( ps \/ ch )
                ( ph -> ch ) ) )
+ bicomi     |- ( ( ( ph -> ps ) \/ ( ph -> ch ) ) <->
                ( ph -> ( ps \/ ch ) ) )
```

```
# (tableau starting atom)
RelStr(c1)
upper(c1)
Subset(union(c2), carrier(c1))
Subset(c2, powerset(carrier(c1)))
```

Figure 2: The MCTS tree for the `WAYBEL_0:28` problem at the moment when the proof is found. For each node we display the predicted probability p, the number of visits n and the average reward $r = w/n$. For the (thicker) nodes leading to the proof the corresponding local proof goals are presented on the right.

Kaliszyk et al 2018

$$3xy\cos(x) - \sqrt{9x^2\sin(x)^2 + 1}y' + 3y\sin(x) = 0$$

$$4x^4yy'' - 8x^4y'^2 - 8x^3yy' - 3x^3y'' - 8x^2y^2 - 6x^2y' - 3x^2y'' - 9xy' - 3y = 0$$

Proof 1

The inversion mapping on G is the mapping $\iota : G \to G$ defined by:

$$\forall g \in G : \iota(g) = g^{-1}$$

where g^{-1} is the inverse or g.

By Inversion Mapping is Involution, ι is an involution:

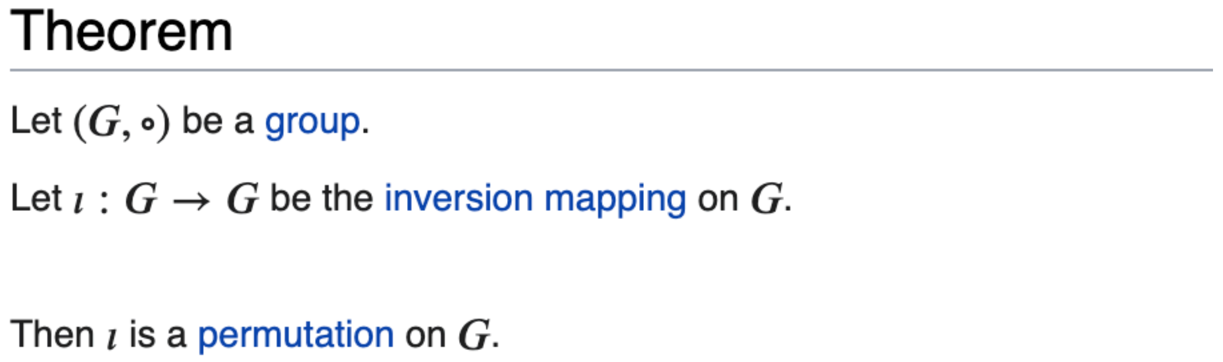
$$\forall g \in G : \iota(\iota(g)) = g$$

The result follows from Involution is Permutation.

■

NaturalProofs tasks

Mathematical reasoning



Theorem

Let $(G, \circ)$ be a group.

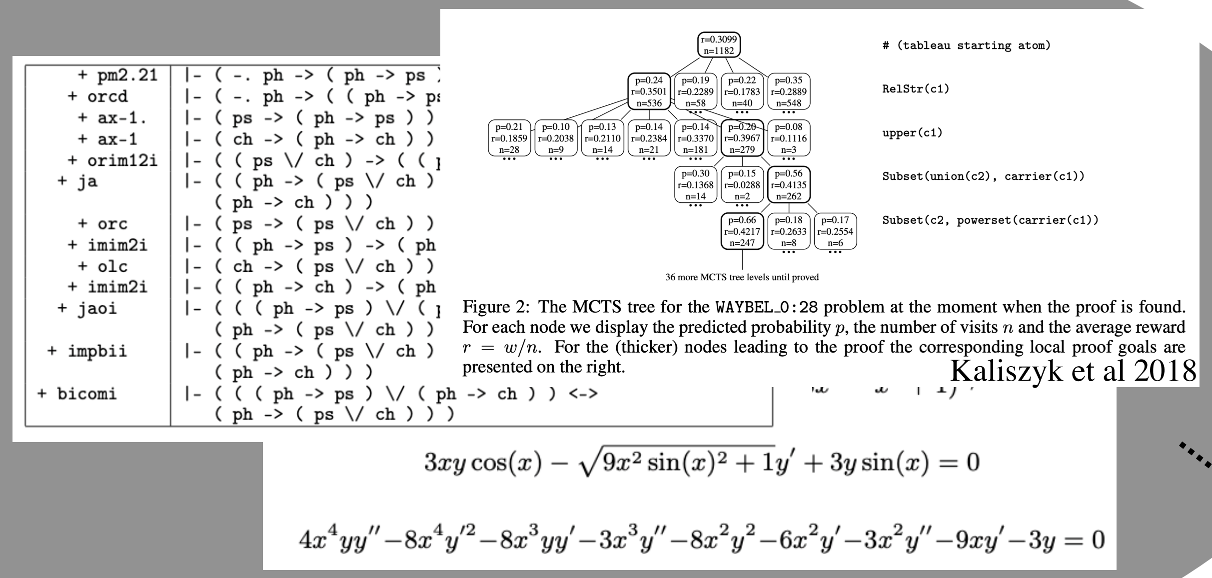
Let $\iota : G \to G$ be the inversion mapping on G.

Then ι is a permutation on G.

Mathematical reasoner

David Hilbert

Symbolic, search



```
  + pm2.21   |- ( -. ph -> ( ph -> ps
  + orcd     |- ( -. ph -> ( ( ph -> ps
  + ax-1.    |- ( ps -> ( ph -> ps ) )
  + ax-1     |- ( ch -> ( ph -> ch ) )
  + orim12i  |- ( ( ps \/ ch ) -> ( (
 + ja        |- ( ( ph -> ( ps \/ ch )
                ( ph -> ch ) ) )
  + orc      |- ( ps -> ( ps \/ ch ) )
 + imim2i    |- ( ( ph -> ps ) -> ( ph
  + olc      |- ( ch -> ( ps \/ ch ) )
 + imim2i    |- ( ( ph -> ch ) -> ( ph
 + jaoi      |- ( ( ( ph -> ps ) \/ (
                ( ph -> ( ps \/ ch ) )
+ impbii     |- ( ( ph -> ( ps \/ ch )
                ( ph -> ch ) ) )
+ bicomi     |- ( ( ( ph -> ps ) \/ ( ph -> ch ) ) <->
                ( ph -> ( ps \/ ch ) ) )
```

Figure 2: The MCTS tree for the `WAYBEL_0:28` problem at the moment when the proof is found. For each node we display the predicted probability p, the number of visits n and the average reward $r = w/n$. For the (thicker) nodes leading to the proof the corresponding local proof goals are presented on the right.

Kaliszyk et al 2018

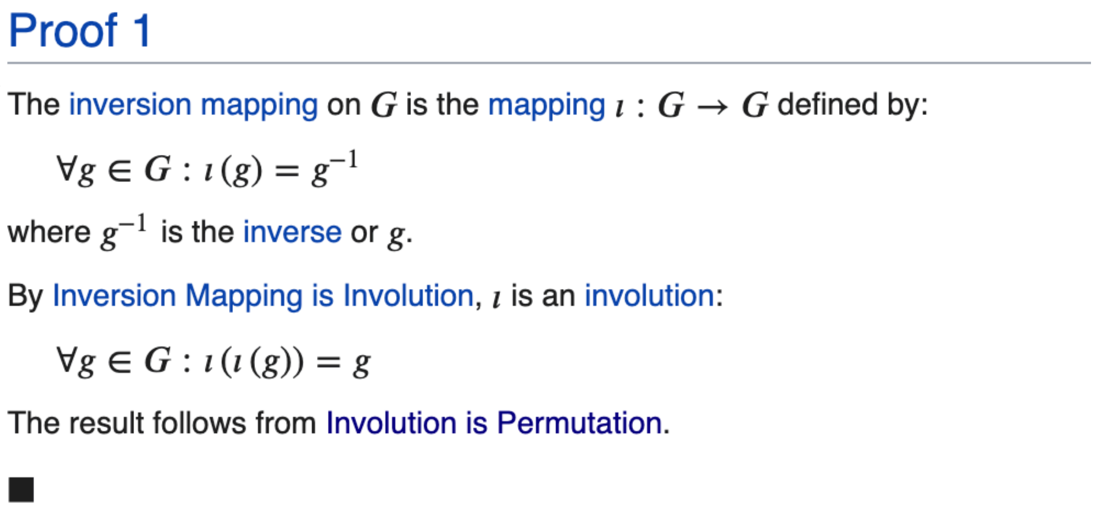
$$3xy\cos(x) - \sqrt{9x^2\sin(x)^2 + 1}y' + 3y\sin(x) = 0$$

$$4x^4yy'' - 8x^4y'^2 - 8x^3yy' - 3x^3y'' - 8x^2y^2 - 6x^2y' - 3x^2y'' - 9xy' - 3y = 0$$

Intuition, analogy, pattern matching

"this seems related to involutions"

"a few days ago I proved that an involution is a permutation…"



Proof 1

The inversion mapping on G is the mapping $\iota : G \to G$ defined by:

$$\forall g \in G : \iota(g) = g^{-1}$$

where g^{-1} is the inverse or g.

By Inversion Mapping is Involution, ι is an involution:

$$\forall g \in G : \iota(\iota(g)) = g$$

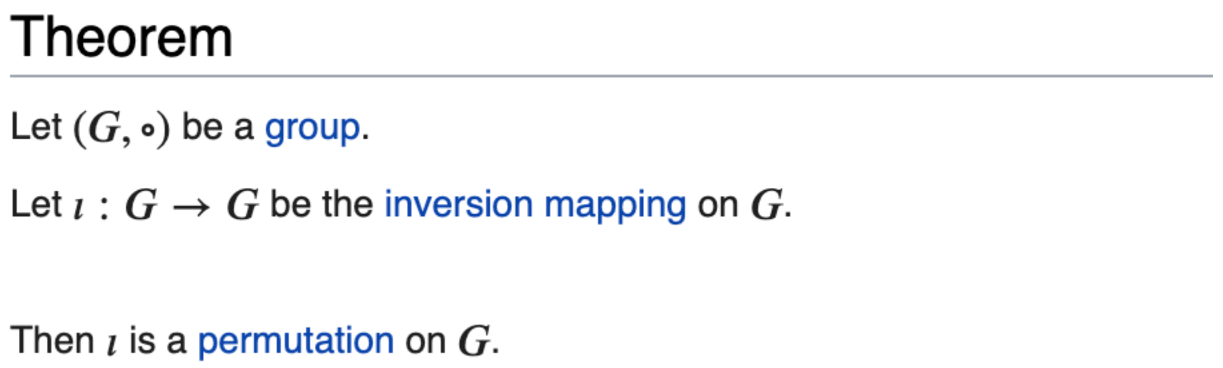
The result follows from Involution is Permutation.

■

"Reference retrieval"
Premise selection

NaturalProofs tasks

Mathematical reasoning



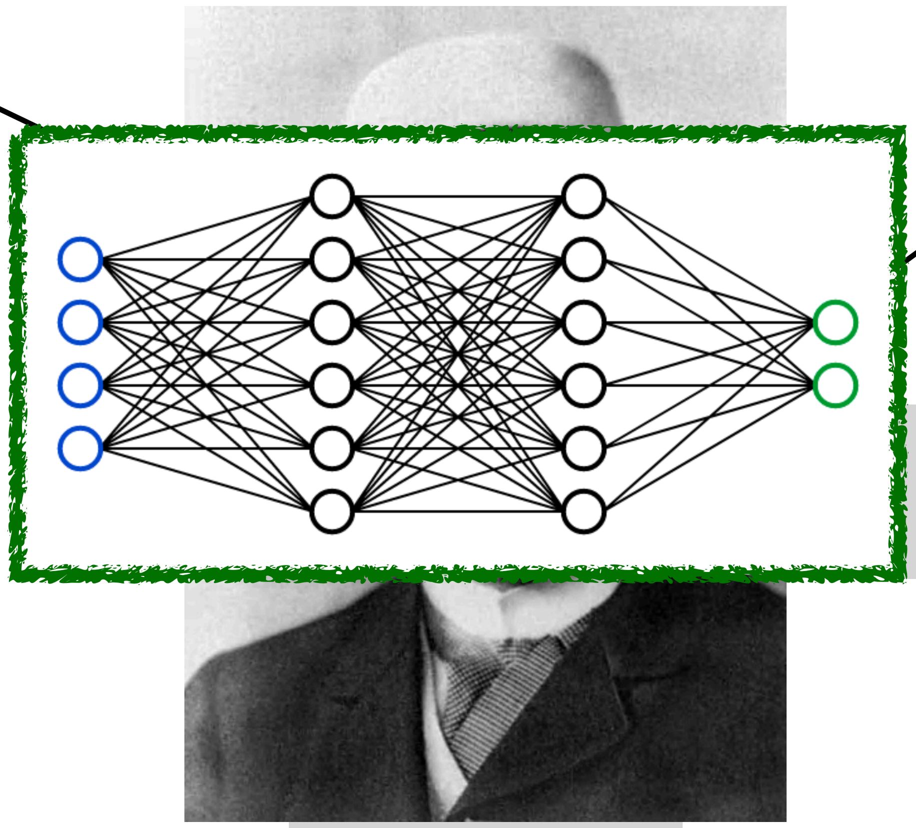
Theorem

Let $(G, \circ)$ be a group.

Let $\iota : G \to G$ be the inversion mapping on G.

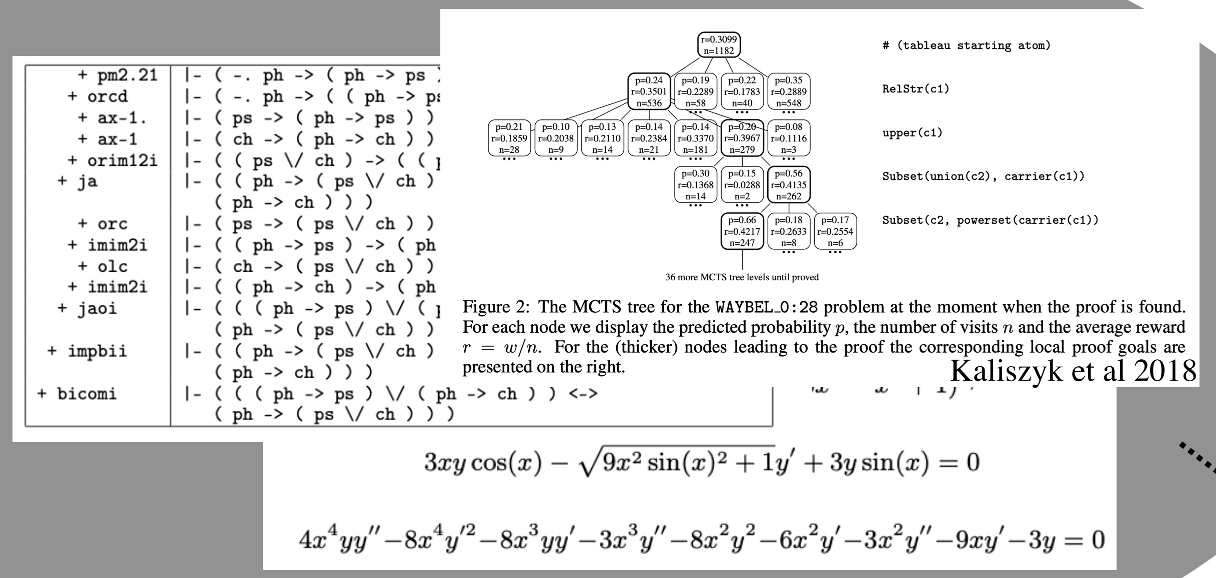
Then ι is a permutation on G.

Mathematical reasoner



David Hilbert

Symbolic, search



```
 + pm2.21  |- ( -. ph -> ( ph -> ps
 + orcd    |- ( -. ph -> ( ( ph -> ps
 + ax-1.   |- ( ps -> ( ph -> ps ) )
 + ax-1    |- ( ch -> ( ph -> ch ) )
 + orim12i |- ( ( ps \/ ch ) -> ( ( |
 + ja      |- ( ( ph -> ( ps \/ ch )
              ( ph -> ch ) ) )
 + orc     |- ( ps -> ( ps \/ ch ) )
 + imim2i  |- ( ( ph -> ps ) -> ( ph
 + olc     |- ( ch -> ( ps \/ ch ) )
 + imim2i  |- ( ( ph -> ch ) -> ( ph
 + jaoi    |- ( ( ( ph -> ps ) \/ ( |
              ( ph -> ( ps \/ ch ) )
 + impbii  |- ( ( ph -> ( ps \/ ch )
              ( ph -> ch ) ) )
 + bicomi  |- ( ( ( ph -> ps ) \/ ( ph -> ch ) ) <->
              ( ph -> ( ps \/ ch ) ) )
```

Figure 2: The MCTS tree for the `WAYBEL_0:28` problem at the moment when the proof is found. For each node we display the predicted probability p, the number of visits n and the average reward $r = w/n$. For the (thicker) nodes leading to the proof the corresponding local proof goals are presented on the right.

Kaliszyk et al 2018

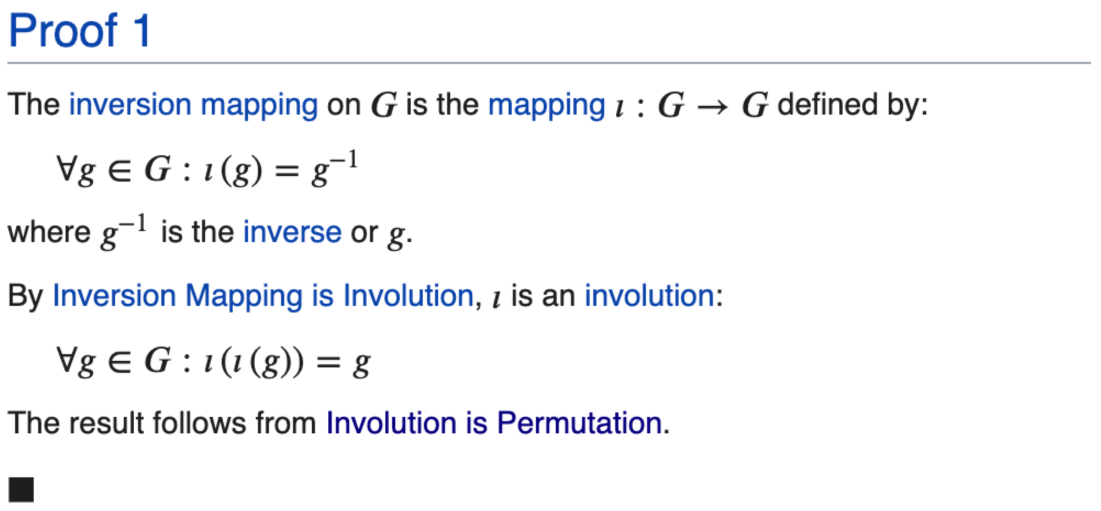
$$3xy\cos(x) - \sqrt{9x^2\sin(x)^2 + 1}y' + 3y\sin(x) = 0$$

$$4x^4yy'' - 8x^4y'^2 - 8x^3yy' - 3x^3y'' - 8x^2y^2 - 6x^2y' - 3x^2y'' - 9xy' - 3y = 0$$

Intuition, analogy, pattern matching

"this seems related to involutions"

"a few days ago I proved that an involution is a permutation…"



Proof 1

The inversion mapping on G is the mapping $\iota : G \to G$ defined by:

$$\forall g \in G : \iota(g) = g^{-1}$$

where g^{-1} is the inverse or g.

By Inversion Mapping is Involution, ι is an involution:

$$\forall g \in G : \iota(\iota(g)) = g$$

The result follows from Involution is Permutation.

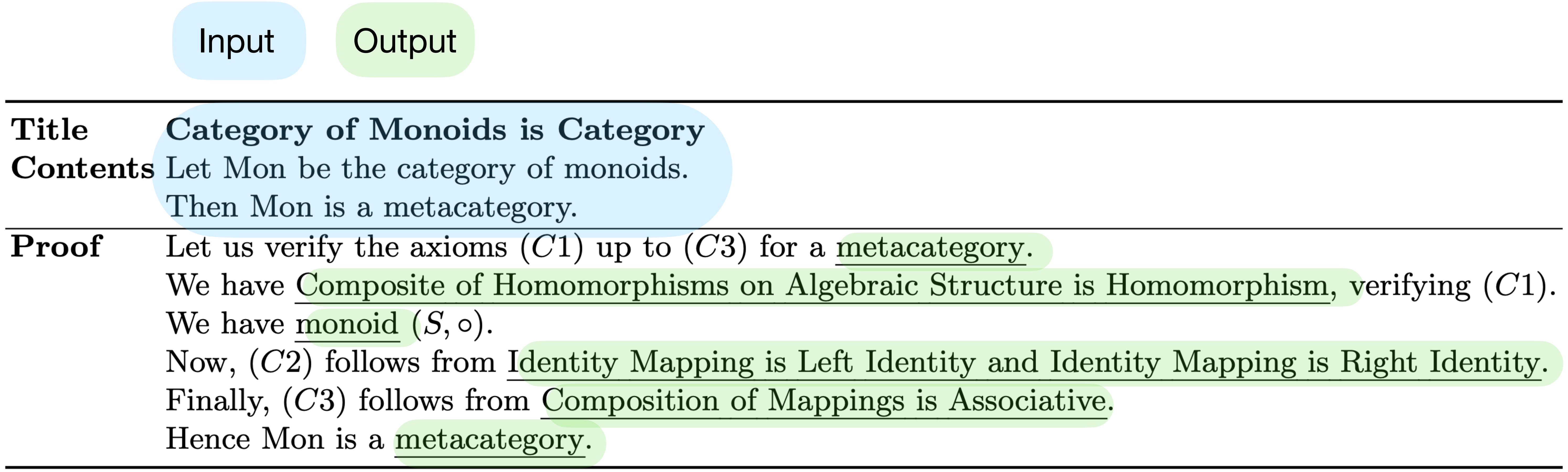
■

"Reference retrieval"
Premise selection

Mathematical reference retrieval

Task

- Given a statement,
  retrieve references that occur in its proof.
- Retrieval:
  - $\mathbf{x}$: theorem
  - $\mathscr{R}$: theorems, definitions, other
  - $\mathbf{r}^{(1)}, \ldots, \mathbf{r}^{(|\mathscr{R}|)}$: ranked list
    - Highly ranked $\Longrightarrow$ in proof of $\mathbf{x}$
- Evaluate with **standard retrieval metrics**

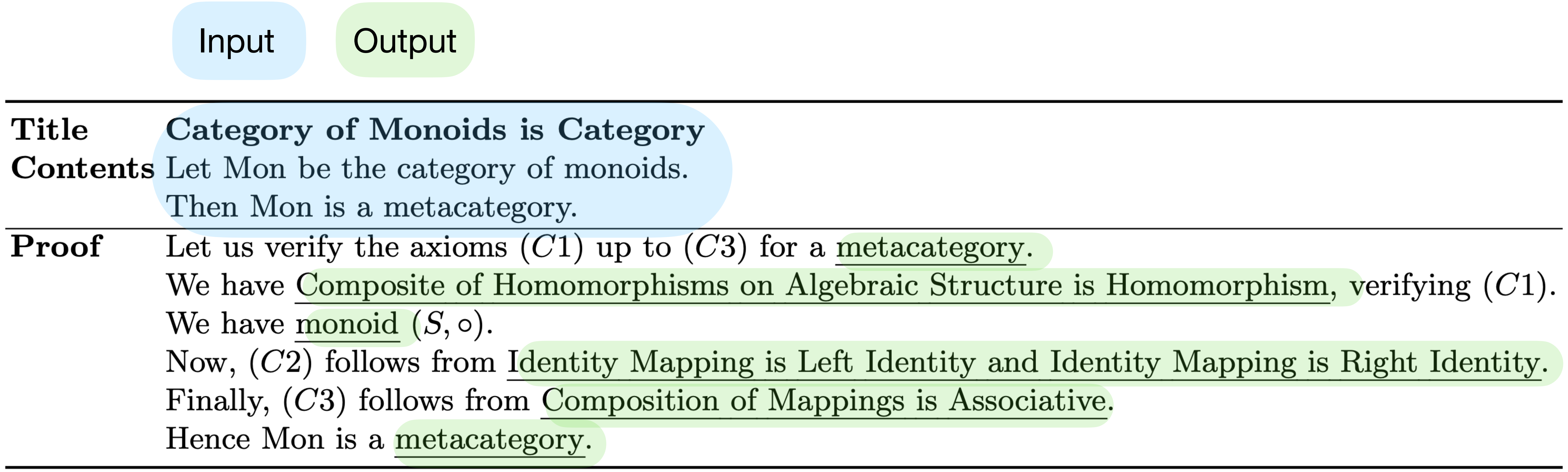


Title	**Category of Monoids is Category**
Contents	Let Mon be the category of monoids. Then Mon is a metacategory.
Proof	Let us verify the axioms $(C1)$ up to $(C3)$ for a metacategory. We have Composite of Homomorphisms on Algebraic Structure is Homomorphism, verifying $(C1)$. We have monoid $(S, \circ)$. Now, $(C2)$ follows from Identity Mapping is Left Identity and Identity Mapping is Right Identity. Finally, $(C3)$ follows from Composition of Mappings is Associative. Hence Mon is a metacategory.

Mathematical reference retrieval

Task

- Given a statement,
  retrieve references that occur in its proof.
- Retrieval:
  - $\mathbf{x}$: theorem
  - $\mathscr{R}$: theorems, definitions, other
  - $\mathbf{r}^{(1)}, \ldots, \mathbf{r}^{(|\mathscr{R}|)}$: ranked list
    - Highly ranked $\Longrightarrow$ in proof of $\mathbf{x}$
- Evaluate with **standard retrieval metrics**



Input Output

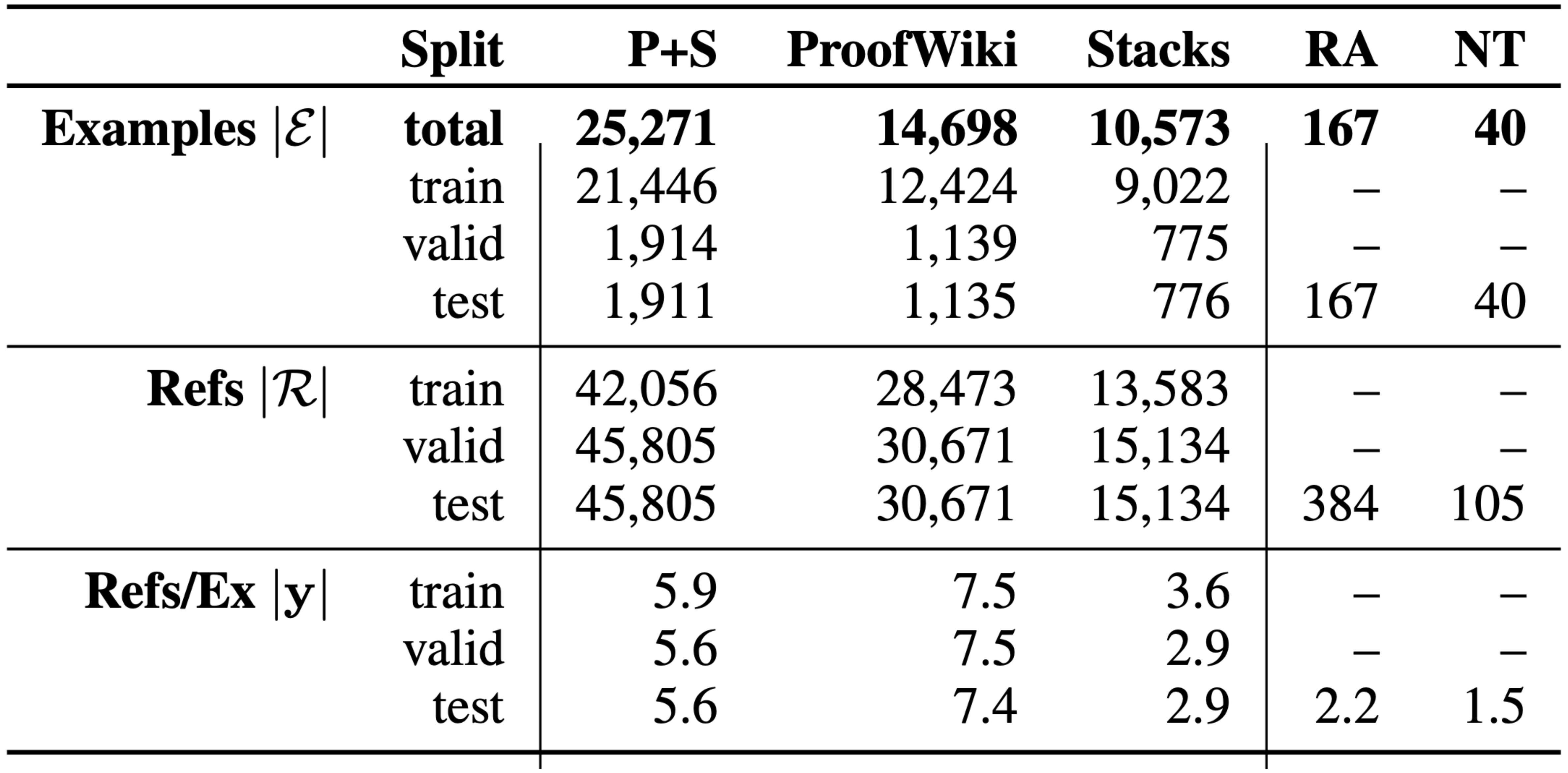
Title	**Category of Monoids is Category**
Contents	Let Mon be the category of monoids. Then Mon is a metacategory.
Proof	Let us verify the axioms $(C1)$ up to $(C3)$ for a metacategory. We have Composite of Homomorphisms on Algebraic Structure is Homomorphism, verifying $(C1)$. We have monoid $(S, \circ)$. Now, $(C2)$ follows from Identity Mapping is Left Identity and Identity Mapping is Right Identity. Finally, $(C3)$ follows from Composition of Mappings is Associative. Hence Mon is a metacategory.

- Premise selection
  - [e.g. Alemi et al 2017, Piotrowski & Urban 2020]
- Natural language premise selection
  - [Ferreira & Freitas 2020 a, b]
    - $\mathbf{r} \in \text{proof}(\mathbf{x})$ for $R \subset \mathscr{R}$ (e.g. $|R| \leq 30$)

Mathematical reference retrieval

Data

- NaturalProofs-derived retrieval dataset
  - 25,000 examples $(\mathbf{x}, \{\mathbf{r}_1, \ldots, \mathbf{r}_{|y|}\})$
  - 45,000 references $|\mathscr{R}|$

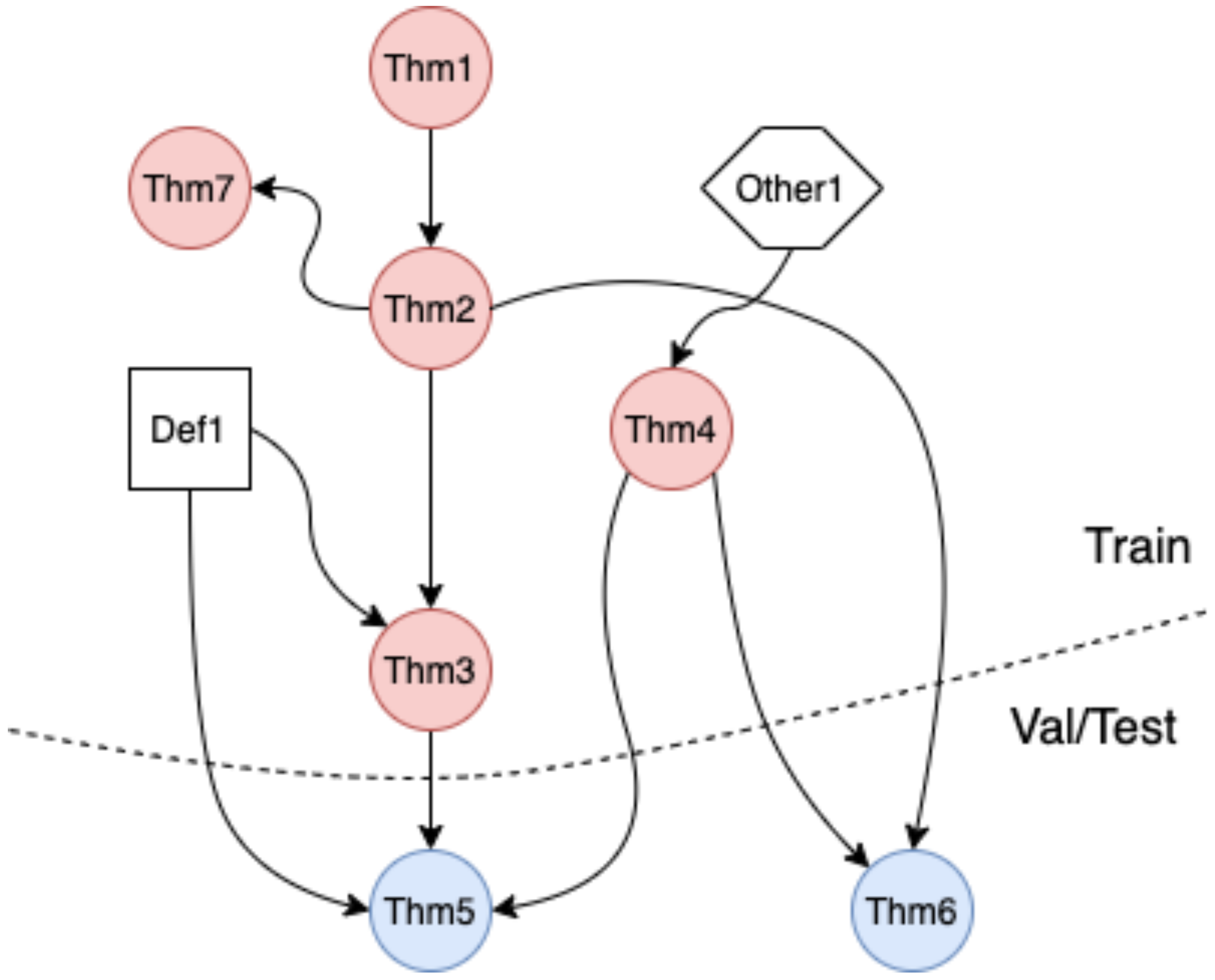


	Split	P+S	ProofWiki	Stacks	RA	NT
Examples $\lvert\mathcal{E}\rvert$	**total**	**25,271**	**14,698**	**10,573**	**167**	**40**
	train	21,446	12,424	9,022	–	–
	valid	1,914	1,139	775	–	–
	test	1,911	1,135	776	167	40
Refs $\lvert\mathcal{R}\rvert$	train	42,056	28,473	13,583	–	–
	valid	45,805	30,671	15,134	–	–
	test	45,805	30,671	15,134	384	105
Refs/Ex $\lvert\mathbf{y}\rvert$	train	5.9	7.5	3.6	–	–
	valid	5.6	7.5	2.9	–	–
	test	5.6	7.4	2.9	2.2	1.5

Mathematical reference retrieval

Data

- NaturalProofs-derived retrieval dataset
  - 25,000 examples $(\mathbf{x}, \{\mathbf{r}_1, \ldots, \mathbf{r}_{|y|}\})$
  - 45,000 references $|\mathscr{R}|$
- Temporal evaluation splits
  - Prove **new** theorems at evaluation time
- Textbooks evaluation set
- References per example:
  - ~7.5 Proofwiki, ~3 Stacks, ~2 textbooks



Mathematical reference retrieval

General objective

- **Learning**: With theorem $\mathbf{x}$, references in proof $\mathbf{y} = \{\mathbf{r}_1, \ldots, \mathbf{r}_{|\mathbf{y}|}\}$
  - True reference distribution $$p_*(\mathbf{r} \,|\, \mathbf{x}) \propto \begin{cases} 1 & \mathbf{r} \in \mathbf{y} \\ 0 & \text{otherwise} \end{cases}$$
  - Goal: match true distribution $$\min_\theta \mathrm{KL}(p_*(\,\cdot\,|\,\mathbf{x}) \,|\, p_\theta(\,\cdot\,|\,\mathbf{x}))$$

$$\equiv \min_\theta -\sum_{\mathbf{r}\in\mathbf{y}} \log \frac{\exp(s_\theta(\mathbf{x}, \mathbf{r}))}{\sum_{\mathbf{r}'\in\mathscr{R}} \exp(s_\theta(\mathbf{x}, \mathbf{r}'))}$$

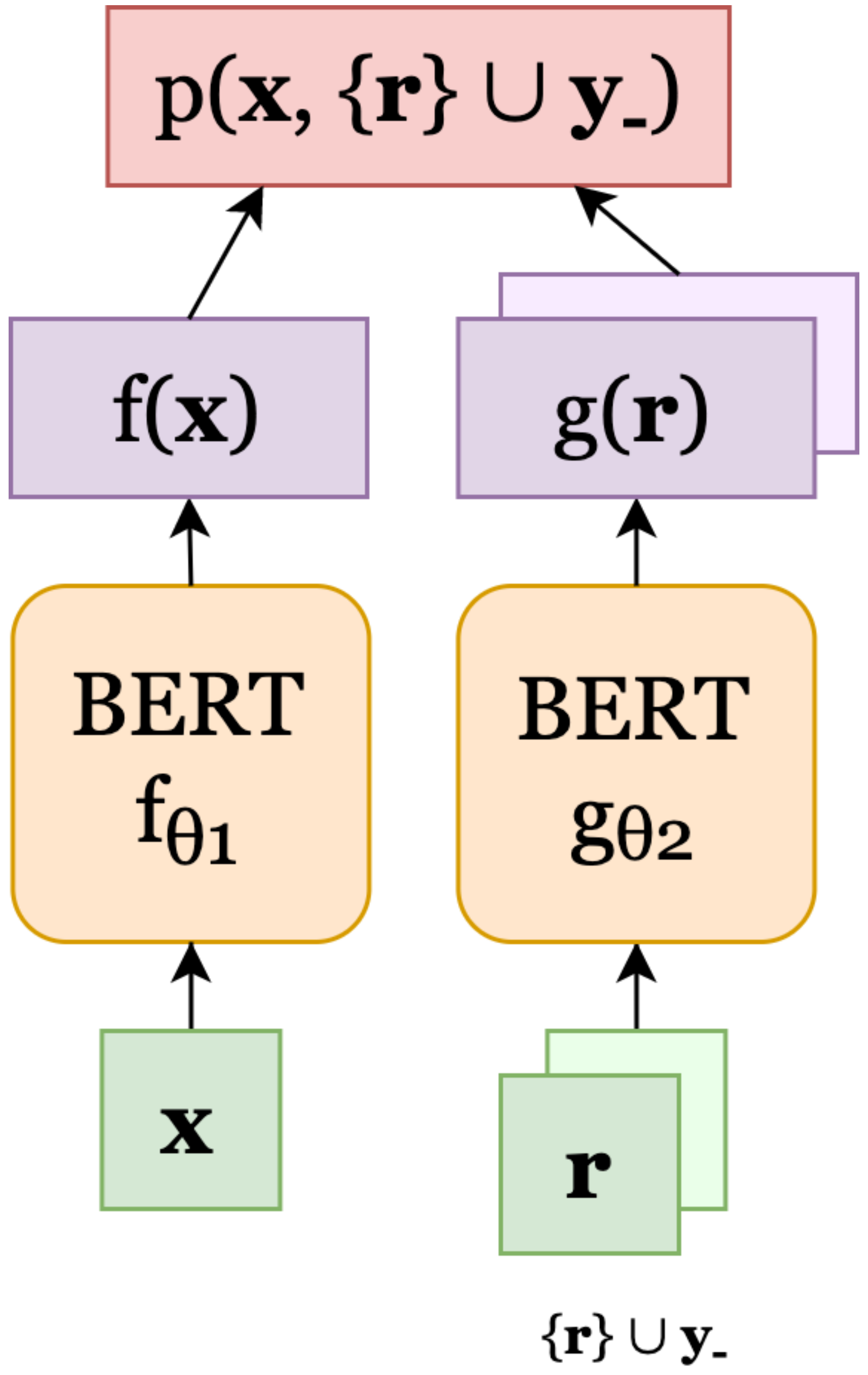
Mathematical reference retrieval

Pairwise model

- **Model 1: “Pairwise"**
  - **Pairwise scoring:** $s_\theta(\mathbf{x}, \mathbf{r}) = f_{\theta_1}(\mathbf{x})^\top g_{\theta_2}(\mathbf{r})$
  - **Approximate loss:** Contrast each reference with negatives

$$\mathscr{L}(\mathbf{x}, \mathbf{r}, \mathbf{y}\_) = -\sum_{\mathbf{r}\in\mathbf{y}} \log \frac{\exp(s_\theta(\mathbf{x}, \mathbf{r}))}{\sum_{\mathbf{r}'\in\mathbf{y}\_} \exp(s_\theta(\mathbf{x}, \mathbf{r}'))}$$

Negatives: other references in the mini-batch
[Karpukhin et al 2020]



Training

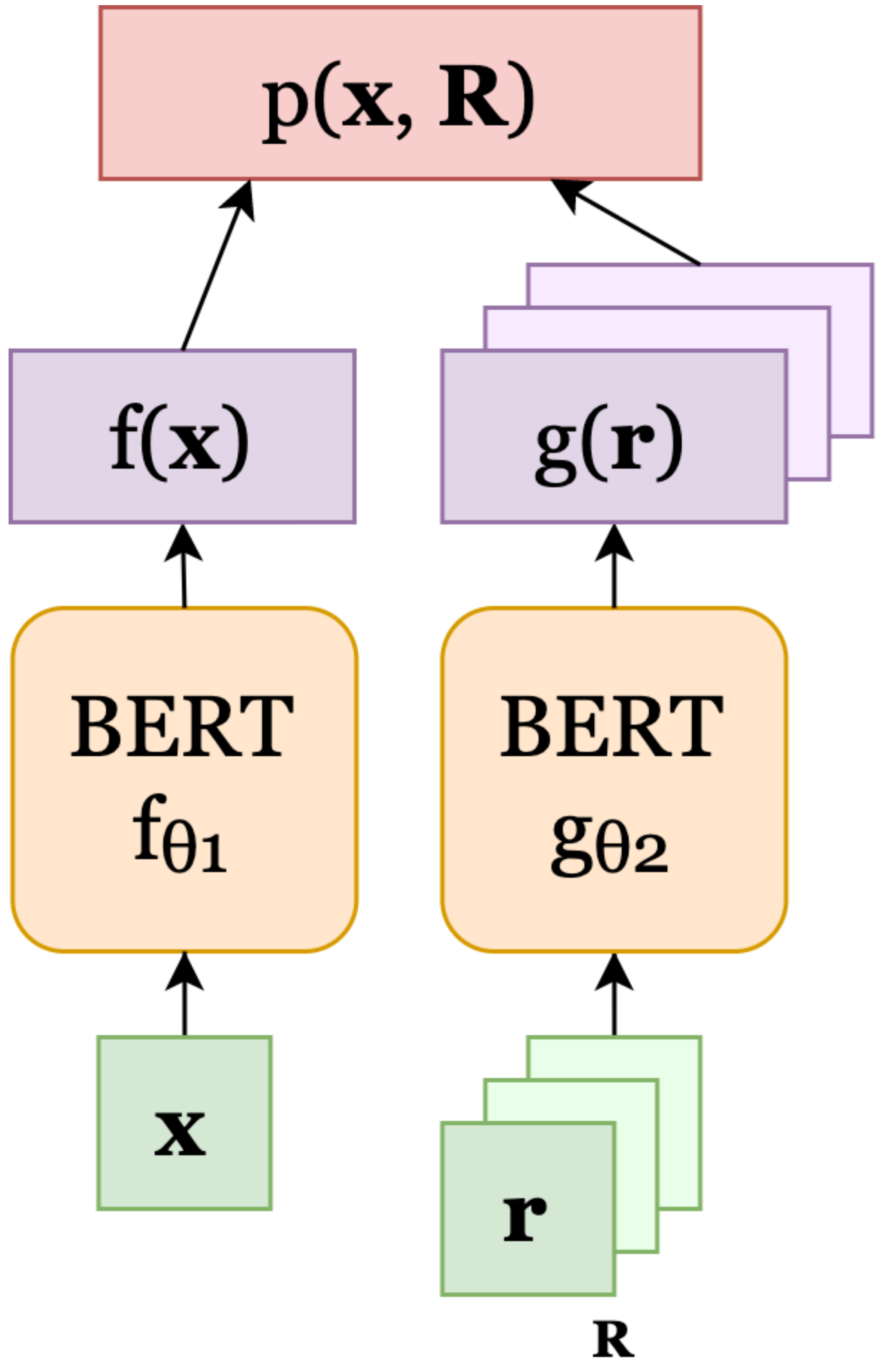
Mathematical reference retrieval

Pairwise model

- **Model 1: "Pairwise"**
  - **Pairwise scoring**: $s_\theta(\mathbf{x}, \mathbf{r}) = f_{\theta_1}(\mathbf{x})^\top g_{\theta_2}(\mathbf{r})$
  - **Approximate loss:** Contrast each reference with negatives

$$\mathscr{L}(\mathbf{x}, \mathbf{r}, \mathbf{y}\_) = -\sum_{\mathbf{r}\in\mathbf{y}} \log \frac{\exp(s_\theta(\mathbf{x}, \mathbf{r}))}{\sum_{\mathbf{r}'\in\mathbf{y}\_} \exp(s_\theta(\mathbf{x}, \mathbf{r}'))}$$

Negatives: other references in the mini-batch
[Karpukhin et al 2020]



Evaluation

Mathematical reference retrieval

Joint model

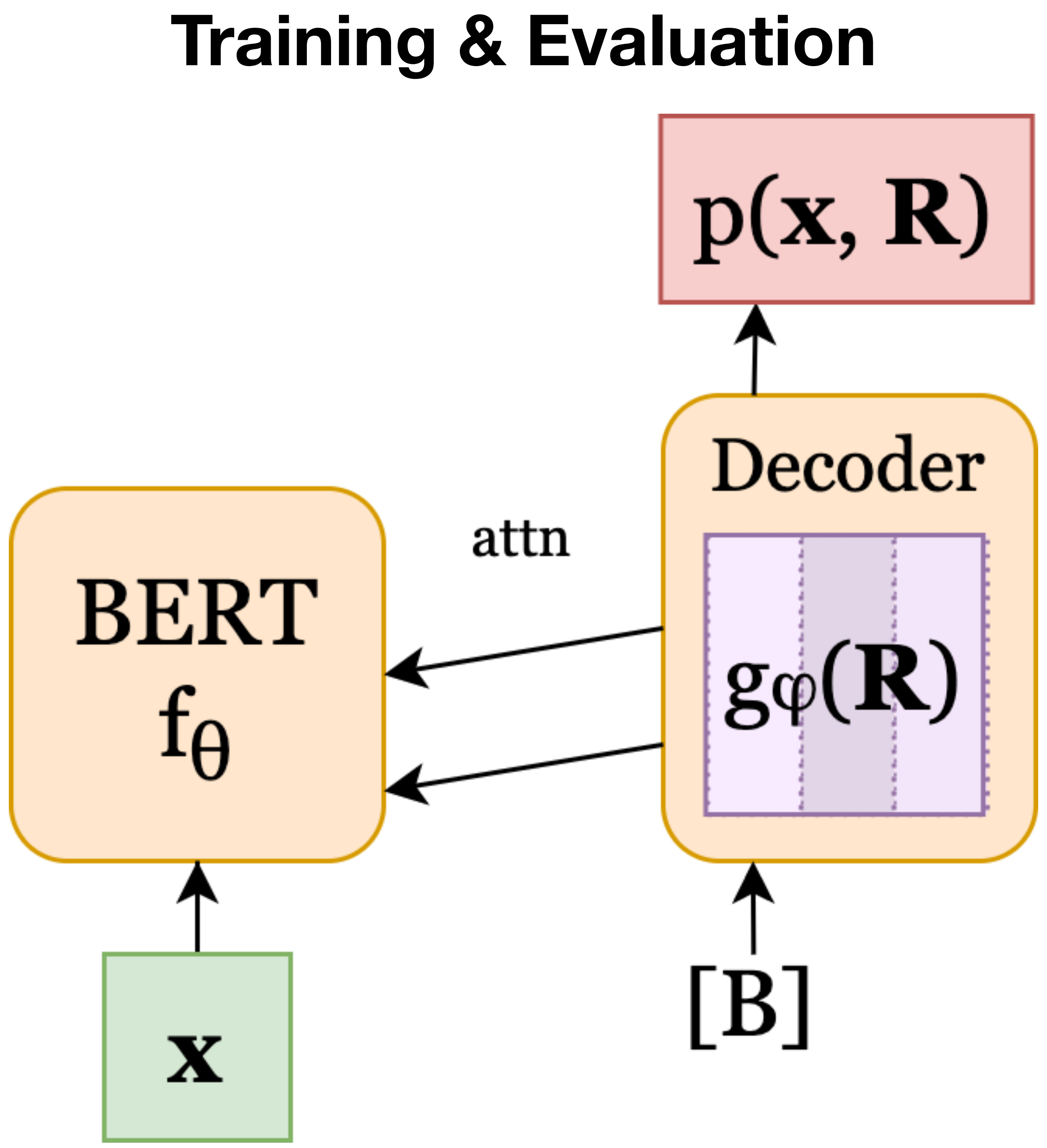
- **Model 2: “Joint”**
  - **Parallel scoring:** $p_\theta(\cdot \mid \mathbf{x}) = \text{softmax}\left(\mathbf{R} f_\theta(\mathbf{x})\right)$

$$\text{where} \quad \mathbf{R} \in \mathbb{R}^{|\mathscr{R}| \times d}$$
$$f_\theta(\mathbf{x}) \in \mathbb{R}^d$$

  - Use the pairwise model’s reference encoder to populate **R**,

$$\mathbf{R} = \begin{bmatrix} -- & g_\phi(\mathbf{r}_1) & -- \\ & \cdots & \\ -- & g_\phi(\mathbf{r}_{|\mathscr{R}|}) & -- \end{bmatrix},$$

  - **Exact loss:** $\mathscr{L}(\mathbf{x}, \mathbf{y}) = \text{KL}(p_*(\cdot \mid \mathbf{x}) \mid p_\theta(\cdot \mid \mathbf{x}))$



Mathematical reference retrieval

Experiments

- **In-domain**
    - Train and evaluate on the same domain
    - Ablations & analysis
- **Out-of-domain**
    - Evaluate on an unseen domain (textbooks)

Mathematical reference retrieval

Experiments | In-domain

		mAP	Recall@10	Full@10	Recall@100	Full@100
PWiki	**TF-IDF**	6.19	10.27	4.14	23.09	9.43
	BERT pair	16.82	23.73	7.31	63.75	38.50
	BERT joint	**36.75**	**42.45**	**20.35**	**75.90**	**50.22**
Stacks	**TF-IDF**	13.64	25.46	18.94	47.36	37.76
	BERT pair	20.93	37.43	30.03	**74.21**	**66.37**
	BERT joint	**28.32**	**39.10**	**31.96**	73.61	65.59

Mathematical reference retrieval

Experiments | In-domain

		mAP	Recall@10	Full@10	Recall@100	Full@100
PWiki	**TF-IDF**	6.19	10.27	4.14	23.09	9.43
	BERT pair	16.82	23.73	7.31	63.75	38.50
	BERT joint	**36.75**	**42.45**	**20.35**	**75.90**	**50.22**
Stacks	**TF-IDF**	13.64	25.46	18.94	47.36	37.76
	BERT pair	20.93	37.43	30.03	**74.21**	**66.37**
	BERT joint	**28.32**	**39.10**	**31.96**	73.61	65.59

✓**BERT** better than classical IR & other baselines

Mathematical reference retrieval

Experiments | In-domain

		mAP	Recall@10	Full@10	Recall@100	Full@100
PWiki	**TF-IDF**	6.19	10.27	4.14	23.09	9.43
	BERT pair	16.82	23.73	7.31	63.75	38.50
	BERT joint	**36.75**	**42.45**	**20.35**	**75.90**	**50.22**
Stacks	**TF-IDF**	13.64	25.46	18.94	47.36	37.76
	BERT pair	20.93	37.43	30.03	**74.21**	**66.37**
	BERT joint	**28.32**	**39.10**	**31.96**	73.61	65.59

✓**BERT** better than classical IR & other baselines

✓**Joint** improves over **pairwise**

Mathematical reference retrieval

Experiments | In-domain

		mAP	Recall@10	Full@10	Recall@100	Full@100
PWiki	**TF-IDF**	6.19	10.27	4.14	23.09	9.43
	BERT pair	16.82	23.73	7.31	63.75	38.50
	BERT joint	**36.75**	**42.45**	**20.35**	**75.90**	**50.22**
Stacks	**TF-IDF**	13.64	25.46	18.94	47.36	37.76
	BERT pair	20.93	37.43	30.03	**74.21**	**66.37**
	BERT joint	**28.32**	**39.10**	**31.96**	73.61	65.59

✓**BERT** better than classical IR & other baselines

✓**Joint** improves over **pairwise**

✓**Top 10** contains ~40% of true references

- **All** true references for 20-30% examples

Mathematical reference retrieval

Experiments | In-domain

		mAP	Recall@10	Full@10	Recall@100	Full@100
PWiki	TF-IDF	6.19	10.27	4.14	23.09	9.43
	BERT pair	16.82	23.73	7.31	63.75	38.50
	BERT joint	**36.75**	**42.45**	**20.35**	**75.90**	**50.22**
Stacks	TF-IDF	13.64	25.46	18.94	47.36	37.76
	BERT pair	20.93	37.43	30.03	**74.21**	**66.37**
	BERT joint	**28.32**	**39.10**	**31.96**	73.61	65.59

✓**BERT** better than classical IR & other baselines

✓**Joint** improves over **pairwise**

✓**Top 10** contains ~40% of true references

- **All** true references for 20-30% examples

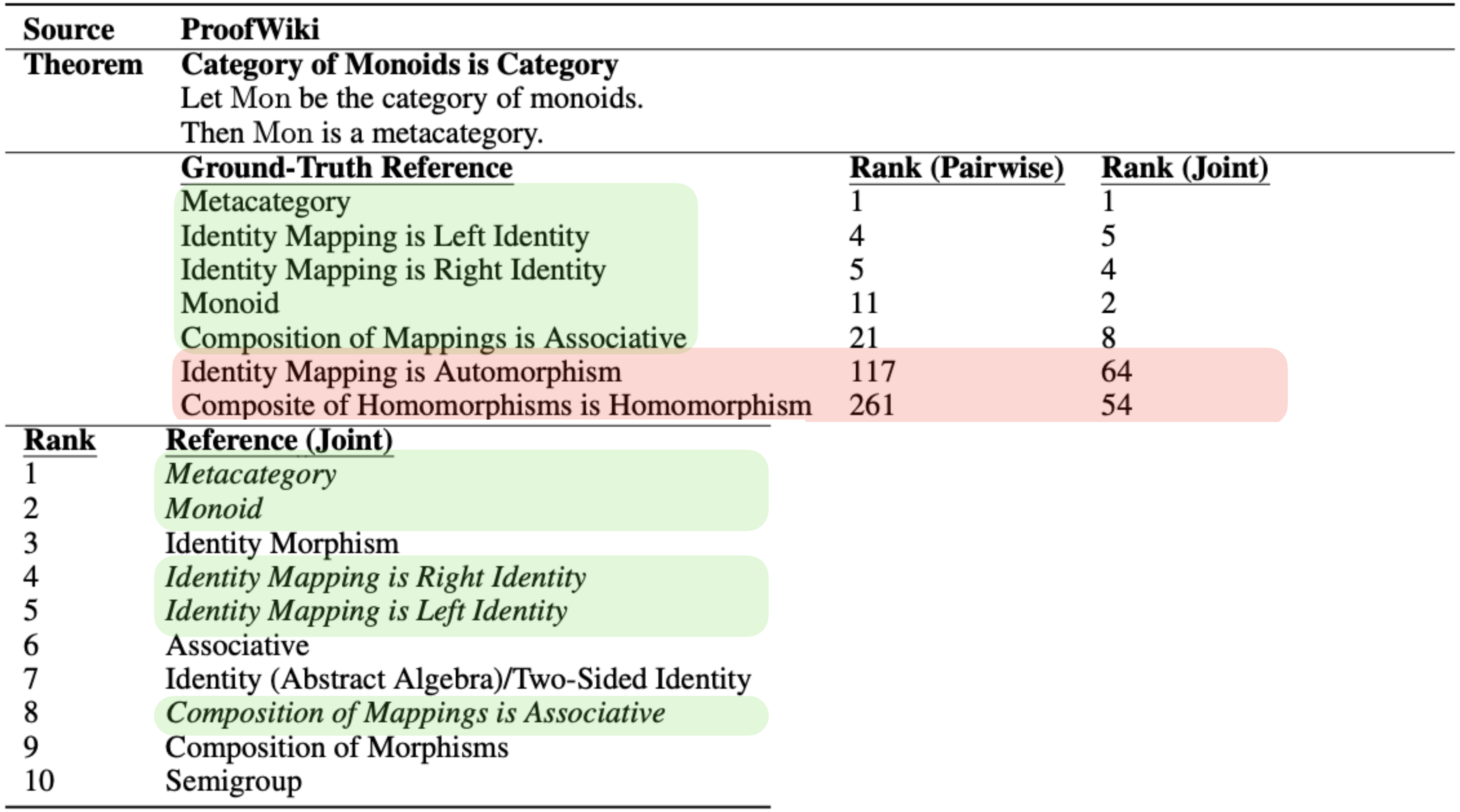
✓**BERT** top 100 contains:

- Roughly **75%** of true references
- **All** true references for ~50-65% of examples

▸ **x** Training on both datasets did not yield improvements

Mathematical reference retrieval

Experiments | Qualitative examples



Source	**ProofWiki**
Theorem	**Category of Monoids is Category** Let Mon be the category of monoids. Then Mon is a metacategory.

True References

Ground-Truth Reference	**Rank (Pairwise)**	**Rank (Joint)**
Metacategory	1	1
Identity Mapping is Left Identity	4	5
Identity Mapping is Right Identity	5	4
Monoid	11	2
Composition of Mappings is Associative	21	8
Identity Mapping is Automorphism	117	64
Composite of Homomorphisms is Homomorphism	261	54

Top-10 model rankings

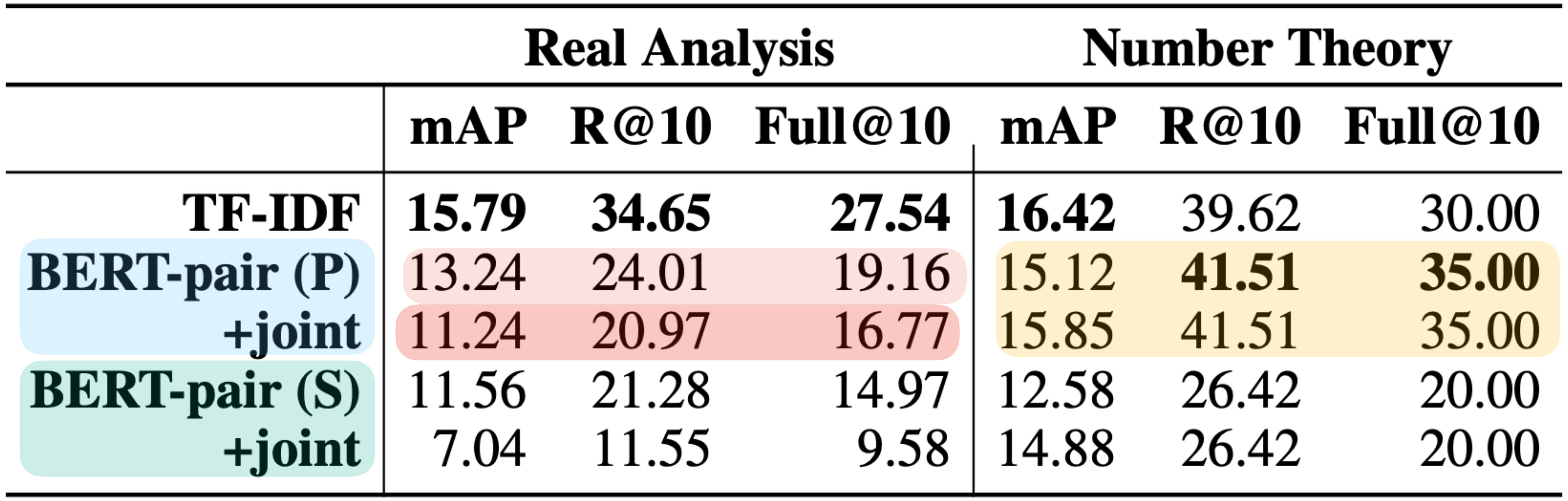
Rank	**Reference (Joint)**
1	*Metacategory*
2	*Monoid*
3	Identity Morphism
4	*Identity Mapping is Right Identity*
5	*Identity Mapping is Left Identity*
6	Associative
7	Identity (Abstract Algebra)/Two-Sided Identity
8	*Composition of Mappings is Associative*
9	Composition of Morphisms
10	Semigroup

Mathematical reference retrieval

Experiments | Out-of-domain

- **x** Neural methods did not generalize well to out-of-domain textbooks
- Training distribution impacts OOD generalization (Proofwiki > Stacks)

Stacks

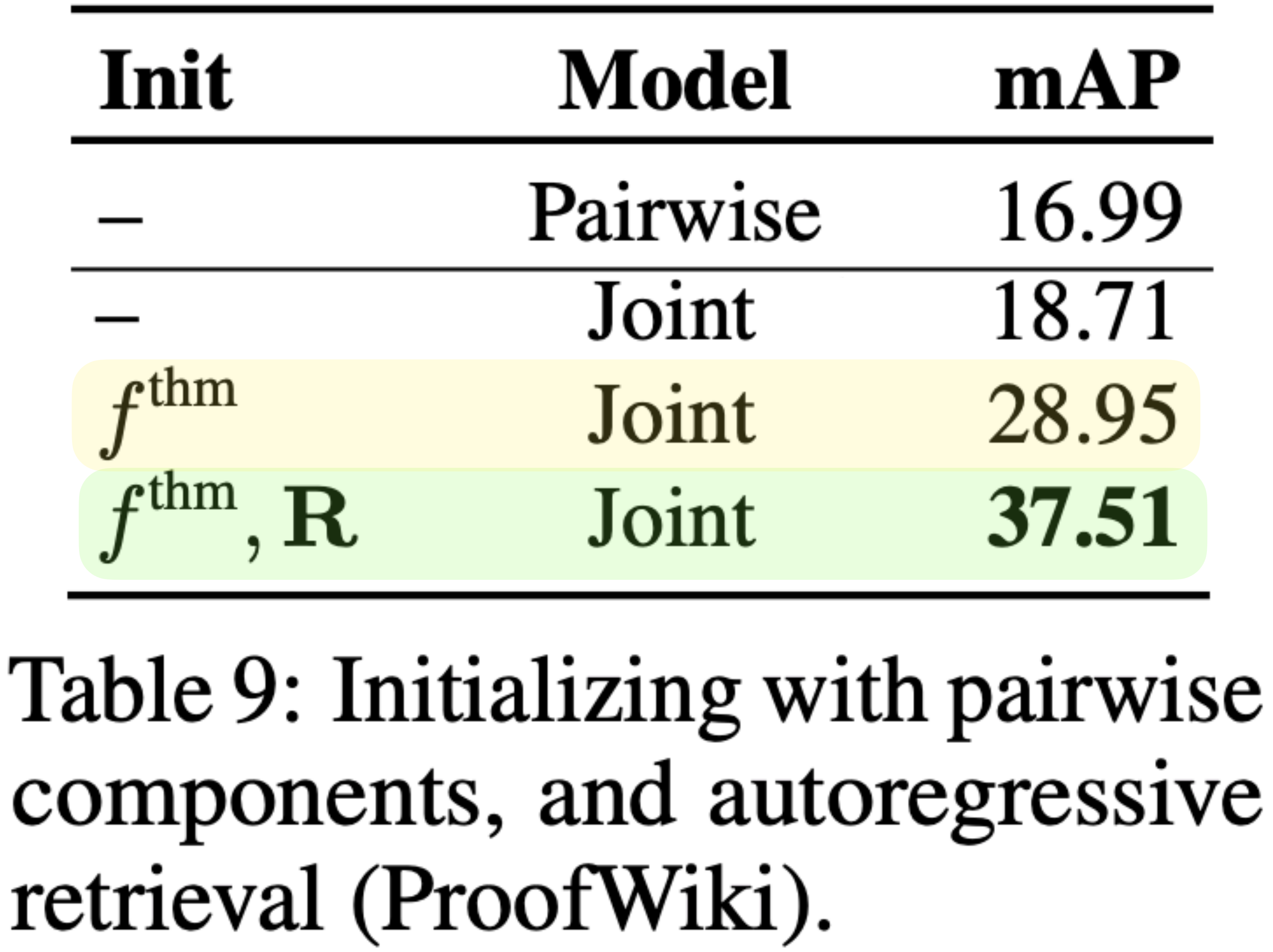


	Real Analysis			Number Theory		
	mAP	**R@10**	**Full@10**	**mAP**	**R@10**	**Full@10**
TF-IDF	**15.79**	**34.65**	**27.54**	**16.42**	39.62	30.00
BERT-pair (P)	13.24	24.01	19.16	15.12	**41.51**	**35.00**
+joint	11.24	20.97	16.77	15.85	41.51	35.00
BERT-pair (S)	11.56	21.28	14.97	12.58	26.42	20.00
+joint	7.04	11.55	9.58	14.88	26.42	20.00

Mathematical reference retrieval

Experiments

- Initializing Joint model with trained pairwise model important
- + using pairwise model's reference embeddings important



Init	Model	mAP
–	Pairwise	16.99
–	Joint	18.71
f^{thm}	Joint	28.95
$f^{\text{thm}}, \mathbf{R}$	Joint	**37.51**

Table 9: Initializing with pairwise components, and autoregressive retrieval (ProofWiki).

Mathematical reference retrieval

Empirical analysis

- Fine-tuning induces semantic groups in reference embeddings

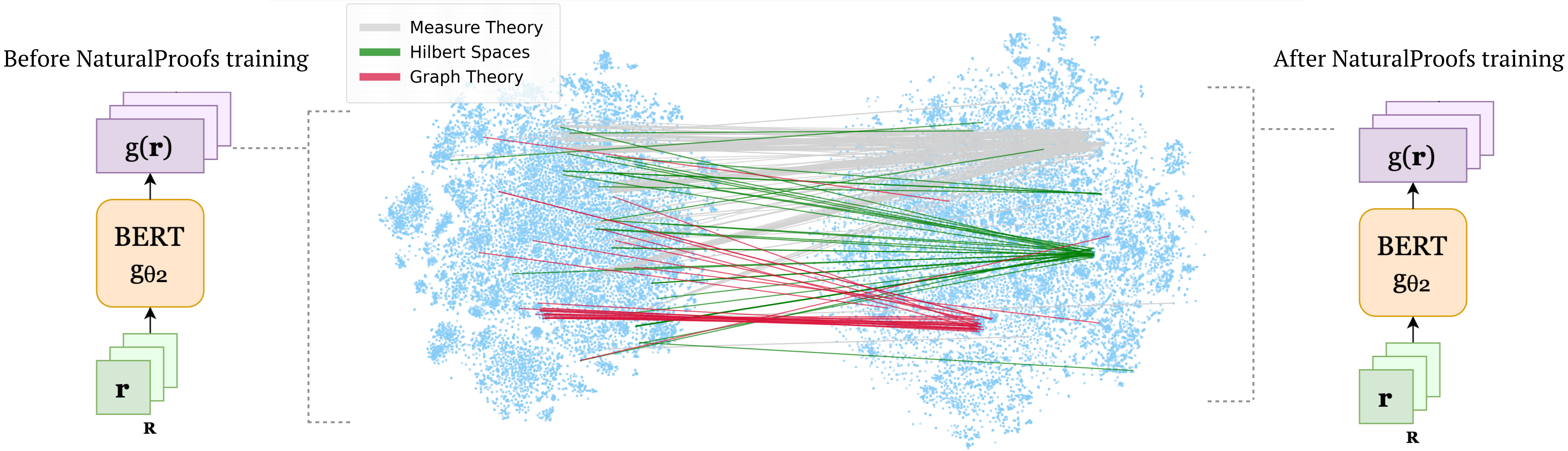


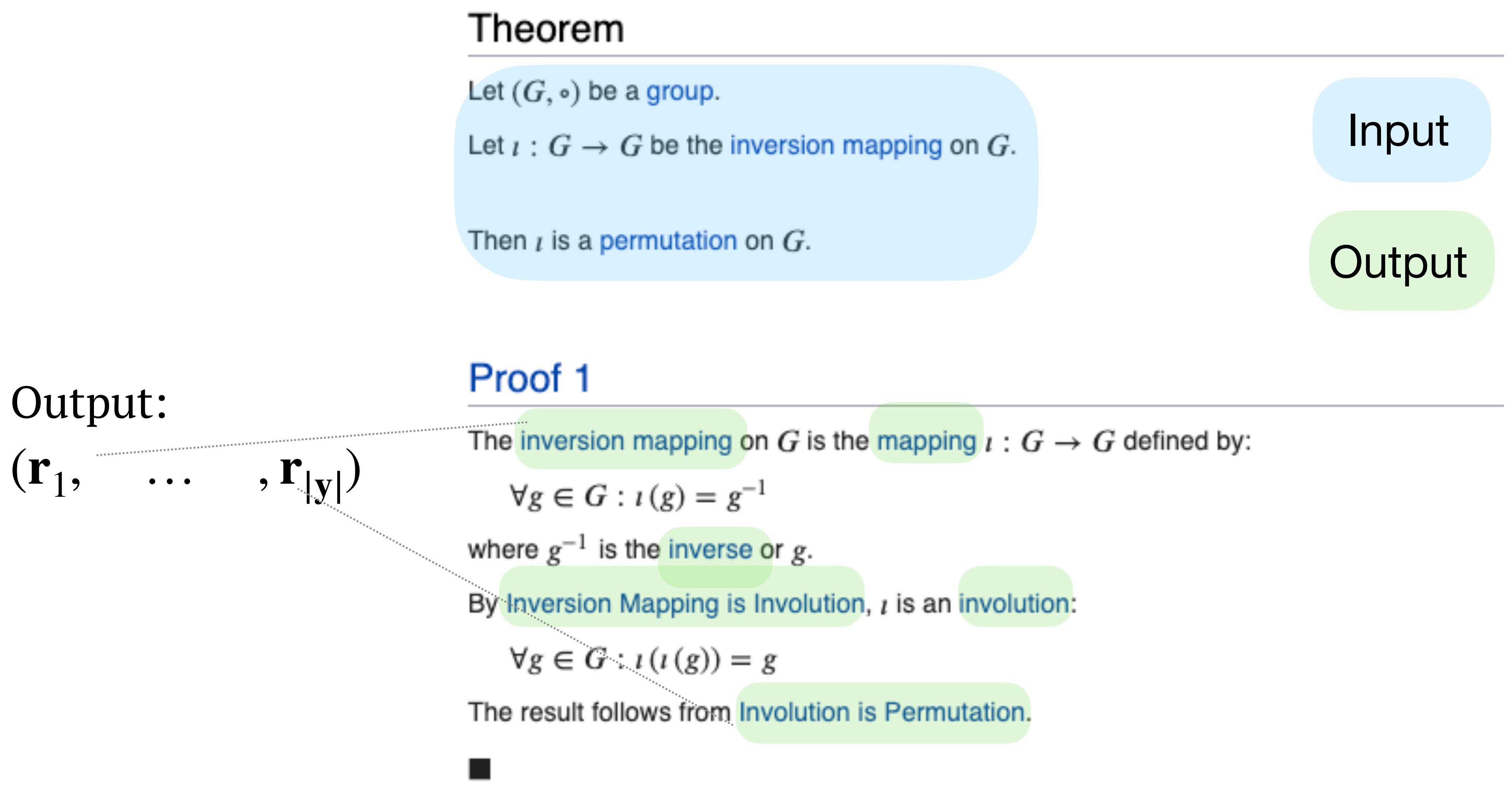
Figure 3. TSNE embeddings for the reference representations before finetuning (left) and after finetuning on NATURALPROOFS (right).

Reference generation

- Given a statement $\mathbf{x}$

- Predict the *sequence* of references in its proof, $\mathbf{y} = (\mathbf{r}_1, \ldots, \mathbf{r}_{|\mathbf{y}|})$

- Autoregressive model (encoder-decoder):
  - $p_\theta(\mathbf{r}_1, \ldots, \mathbf{r}_{|\mathbf{y}|} \mid \mathbf{x}) = \prod_{t=1}^{|\mathbf{y}|+1} p_\theta(\mathbf{r}_t \mid \mathbf{r}_{<t}, \mathbf{x}),$



Output: $(\mathbf{r}_1, \quad \ldots \quad, \mathbf{r}_{|\mathbf{y}|})$

Theorem

Let $(G, \circ)$ be a group.

Let $\iota : G \to G$ be the inversion mapping on G.

Then ι is a permutation on G.

Proof 1

The inversion mapping on G is the mapping $\iota : G \to G$ defined by:

$$\forall g \in G : \iota(g) = g^{-1}$$

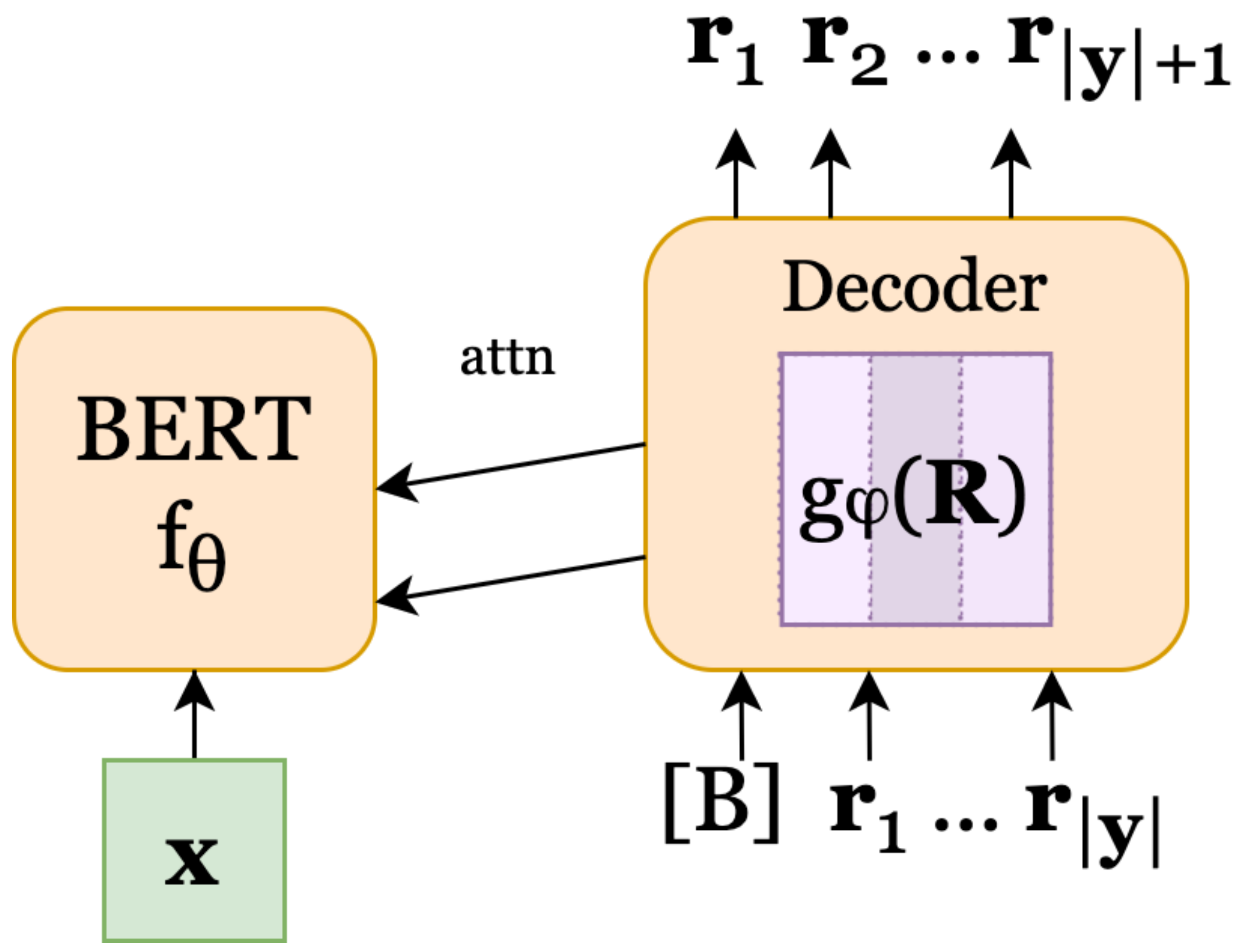
where g^{-1} is the inverse or g.

By Inversion Mapping is Involution, ι is an involution:

$$\forall g \in G : \iota(\iota(g)) = g$$

The result follows from Involution is Permutation.

■

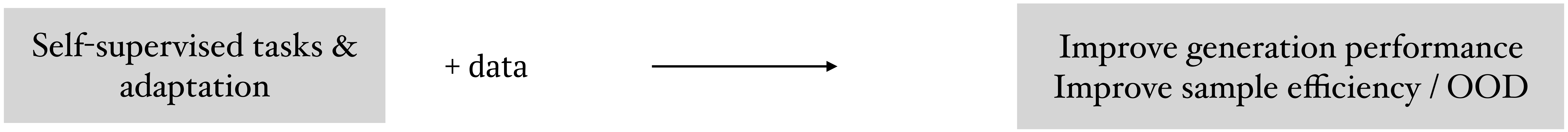


Reference generation

- Oracle benchmarks:
    - Correct *set* (random order)
    - Correct *multiset* (random order)
    - Correct 1st *half of sequence*
- Large room for improvement

	Edit(↓)	**BLEU**(↑)	**EM**(↑)	**F1**(↑)
*-set	58.51	7.18	18.09	97.04
*-multiset	58.09	16.68	19.23	100.0
*-halfseq	58.84	25.88	0.00	56.86
Joint	93.03	0.00	0.09	25.30
Sequential	84.30	5.48	3.78	25.61

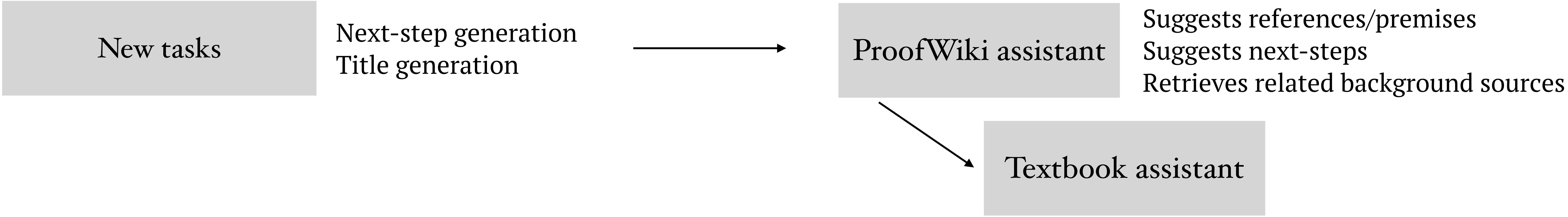
- **Motivation**: "Mathematical assistant"
- **Data**: Multi-domain NaturalProofs
- **Tasks**: Reference retrieval & generation
- **Future directions**


Self-supervised tasks & adaptation

+ data

Improve generation performance
Improve sample efficiency / OOD

- **Domain-specific self-supervised tasks benefit formalized mathematics**
  - e.g. Skip-tree [Rabe et al 2020], LIME [Wu et al 2021], PACT [Han et al 2021]
  - *What are effective self-supervised tasks for informal mathematics?*



New tasks

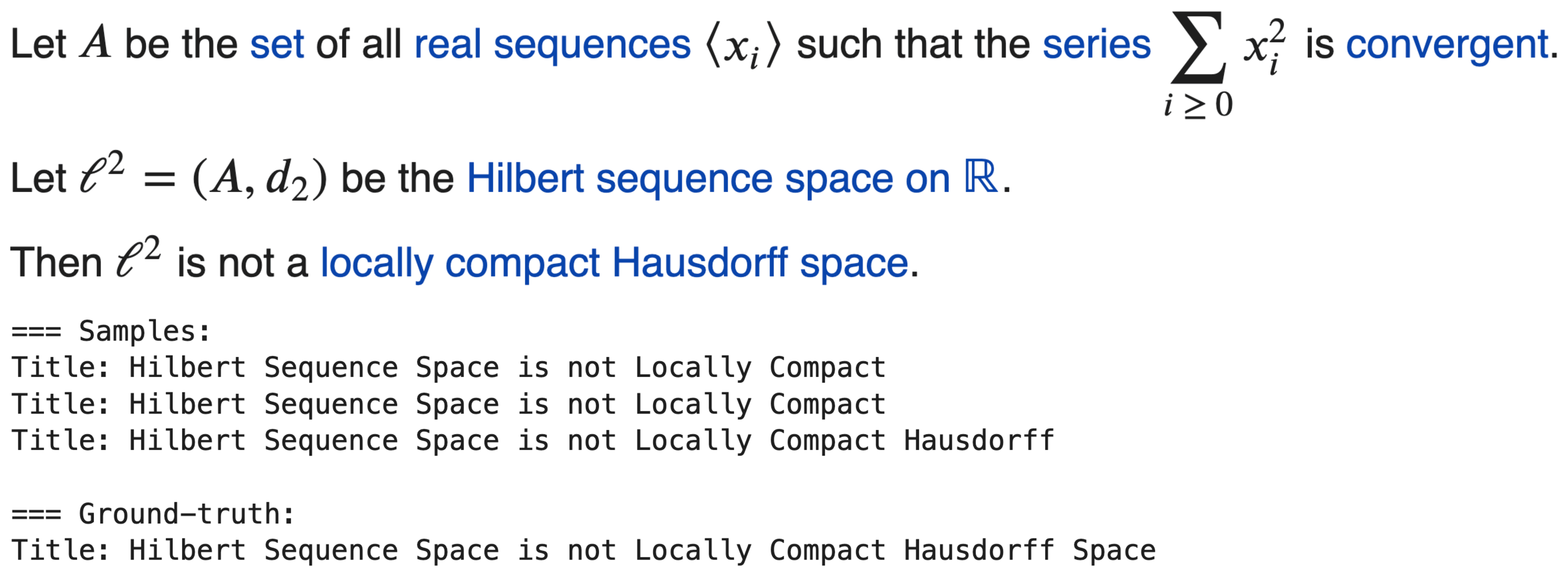
Next-step generation
Title generation

ProofWiki assistant

Suggests references/premises
Suggests next-steps
Retrieves related background sources

Textbook assistant

- **Other tasks with informal mathematics**
- Preliminary exploration:
    - BART title generation



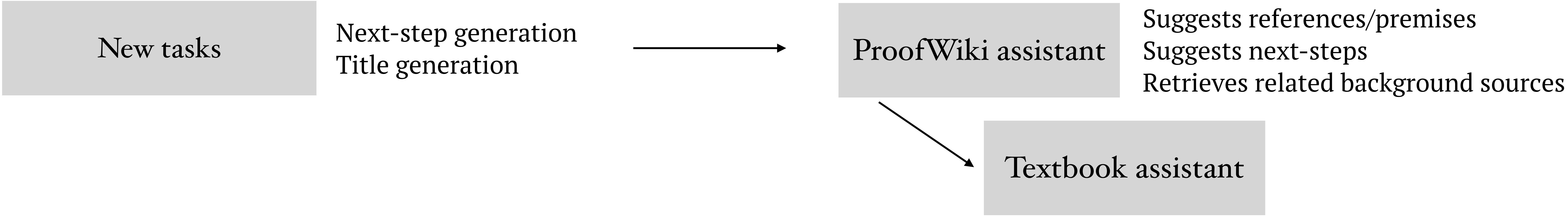
Let A be the set of all real sequences $\langle x_i \rangle$ such that the series $\displaystyle\sum_{i \ge 0} x_i^2$ is convergent.

Let $\ell^2 = (A, d_2)$ be the Hilbert sequence space on $\mathbb{R}$.

Then ℓ^2 is not a locally compact Hausdorff space.

```
=== Samples:
Title: Hilbert Sequence Space is not Locally Compact
Title: Hilbert Sequence Space is not Locally Compact
Title: Hilbert Sequence Space is not Locally Compact Hausdorff

=== Ground-truth:
Title: Hilbert Sequence Space is not Locally Compact Hausdorff Space
```



New tasks

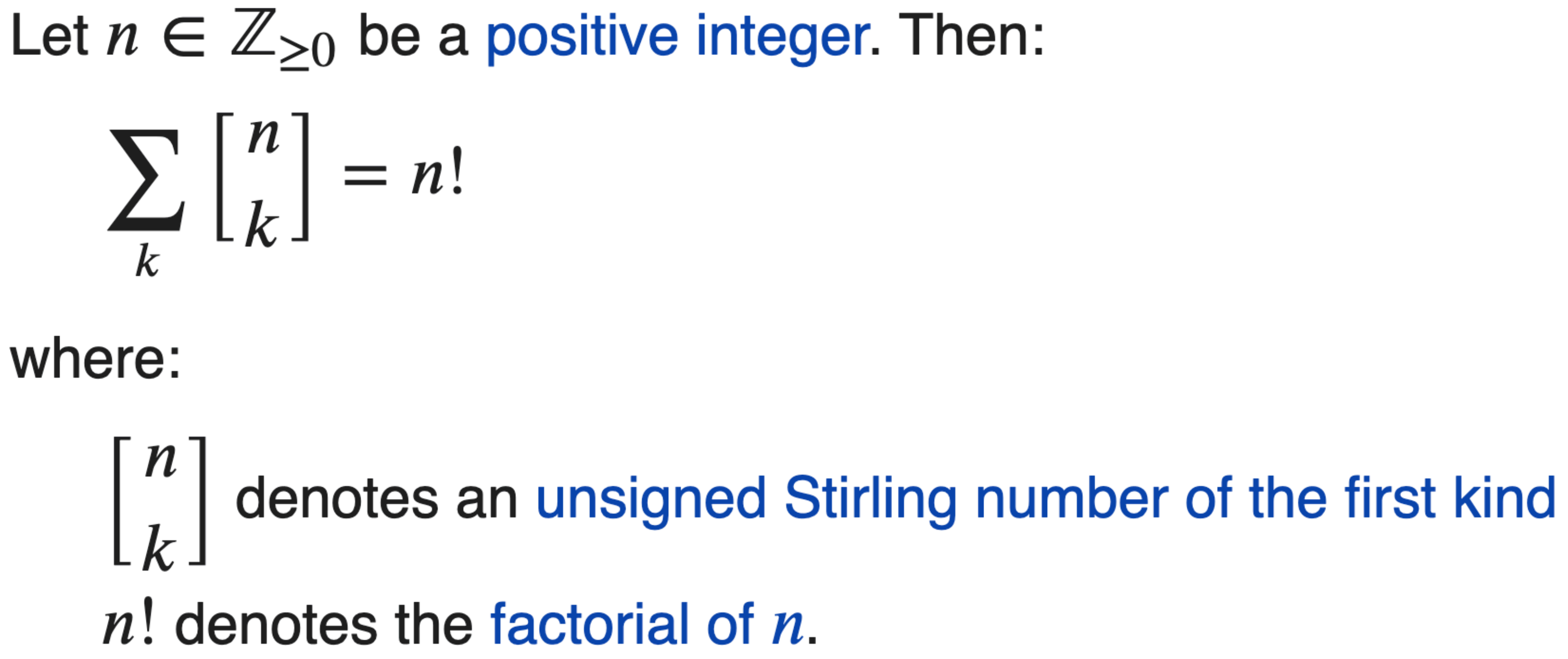
Next-step generation
Title generation

ProofWiki assistant

Suggests references/premises
Suggests next-steps
Retrieves related background sources

Textbook assistant

- **Other tasks with informal mathematics**
- Preliminary exploration:
  - BART title generation
  - Next-step generation
    - Retrieval-augmented generation
- Note: no formal grounding for eval



Let $n \in \mathbb{Z}_{\ge 0}$ be a positive integer. Then:

$$\sum_k \left[{n \atop k}\right] = n!$$

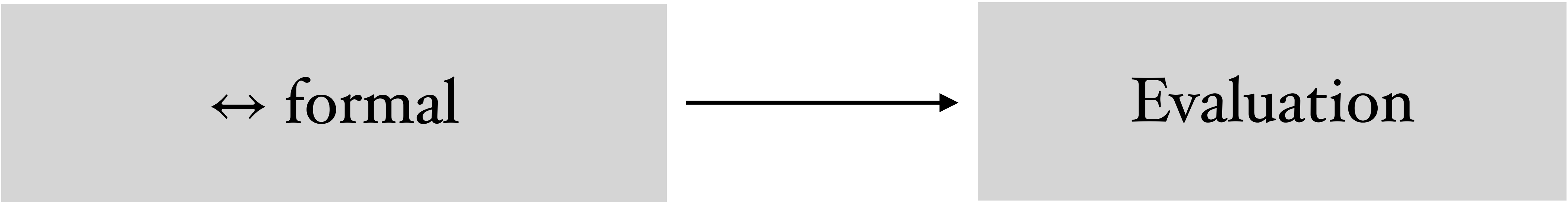
where:

$\left[{n \atop k}\right]$ denotes an unsigned Stirling number of the first kind

$n!$ denotes the factorial of n.

```
=== Samples:
Title: Sum of Unsigned Stirling Numbers of the First Kind
Title: Sum of Stirling Numbers of the First Kind
Title: Sum of Unsigned Stirling Numbers of the First Kind

=== Ground-truth:
Title: Summation over Lower Index of Unsigned Stirling Numbers of the First Kind
```



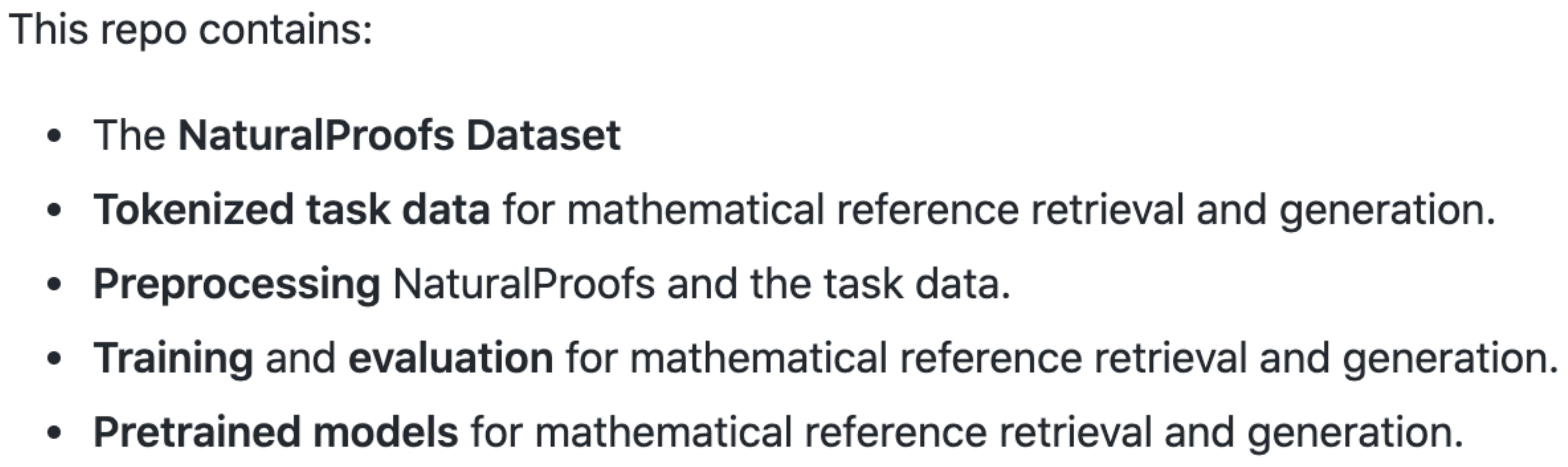
Translation
Aligned bootstrapping
Controlled languages?

- Retrieval-augmented generation
  - Induce association between informal and formal
- Bootstrapping from aligned corpora

Thank you!

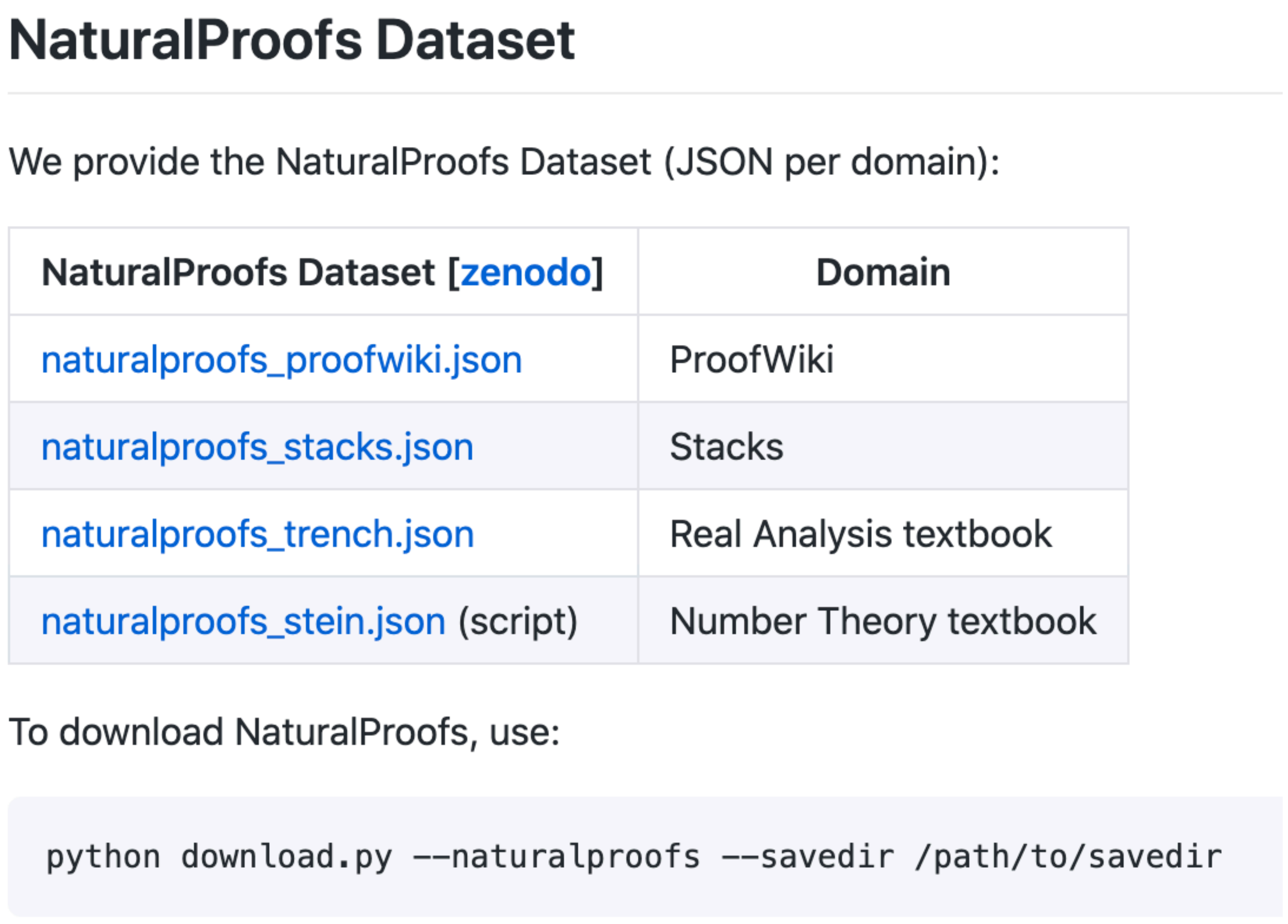
Resources

- Data/models/code: https://github.com/wellecks/naturalproofs



This repo contains:

- The **NaturalProofs Dataset**
- **Tokenized task data** for mathematical reference retrieval and generation.
- **Preprocessing** NaturalProofs and the task data.
- **Training** and **evaluation** for mathematical reference retrieval and generation.
- **Pretrained models** for mathematical reference retrieval and generation.



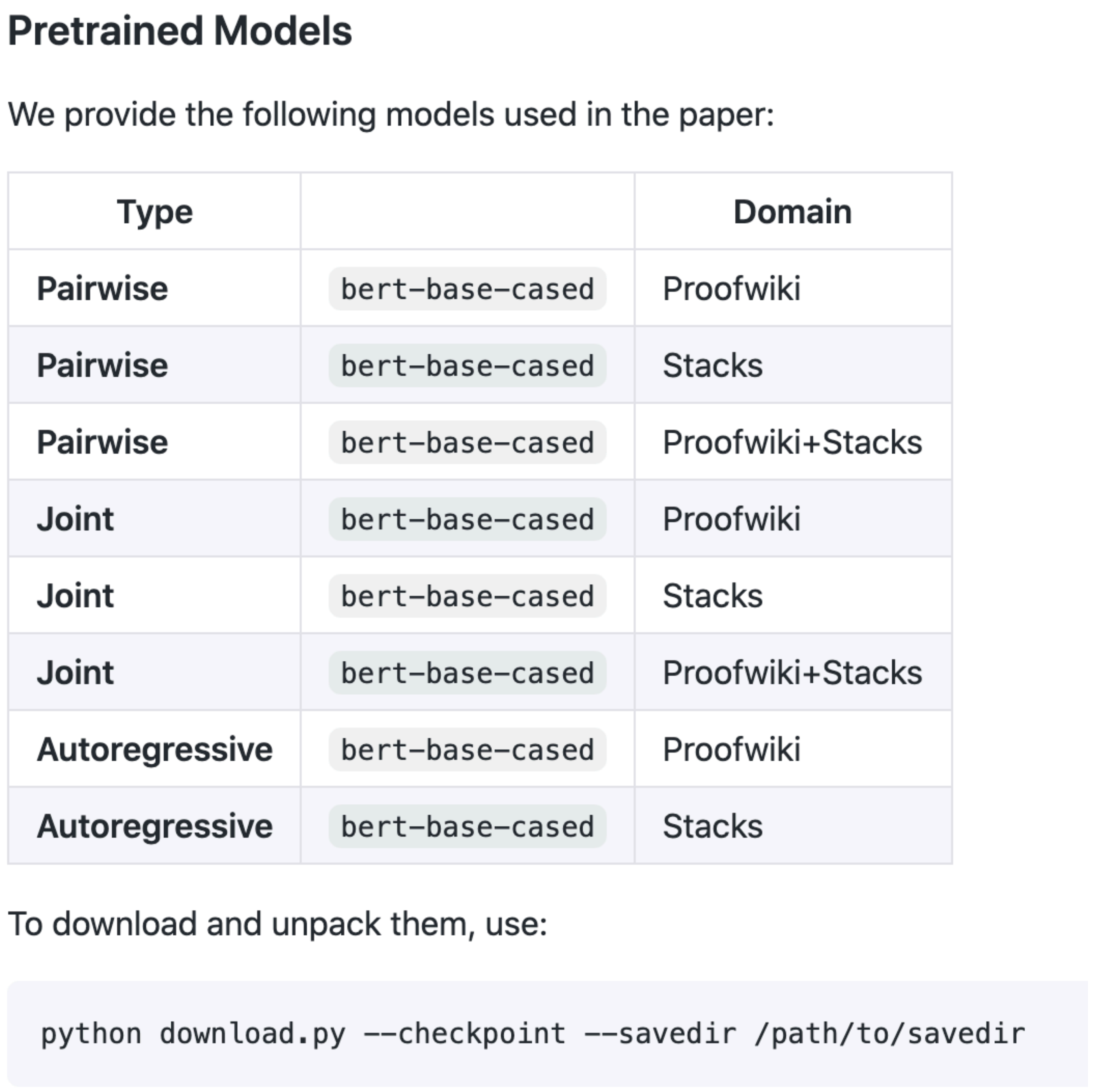
NaturalProofs Dataset

We provide the NaturalProofs Dataset (JSON per domain):

NaturalProofs Dataset [zenodo]	Domain
naturalproofs_proofwiki.json	ProofWiki
naturalproofs_stacks.json	Stacks
naturalproofs_trench.json	Real Analysis textbook
naturalproofs_stein.json (script)	Number Theory textbook

To download NaturalProofs, use:

```
python download.py --naturalproofs --savedir /path/to/savedir
```



Pretrained Models

We provide the following models used in the paper:

Type		Domain
Pairwise	`bert-base-cased`	Proofwiki
Pairwise	`bert-base-cased`	Stacks
Pairwise	`bert-base-cased`	Proofwiki+Stacks
Joint	`bert-base-cased`	Proofwiki
Joint	`bert-base-cased`	Stacks
Joint	`bert-base-cased`	Proofwiki+Stacks
Autoregressive	`bert-base-cased`	Proofwiki
Autoregressive	`bert-base-cased`	Stacks

To download and unpack them, use:

```
python download.py --checkpoint --savedir /path/to/savedir
```

- Neurips 2021 Datasets & Benchmarks: https://arxiv.org/pdf/2104.01112.pdf